

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/373,134D
; FILING DATE: January 17, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Friebel, Thomas E.
; REGISTRATION NUMBER: 29,258
; REFERENCE/DOCKET NUMBER: 7991-007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 781 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-373-134D-2

```

```

Query Match          69.4%; Score 34; DB 1; Length 781;
Best Local Similarity 55.6%; Pred. No. 1.3e+02;
Matches      5; Conservative      2; Mismatches      2; Indels      0; Gaps      0;

```

```

Qy      1 CLSSRLDAC 9
        |:| |: ||
Db      11 CISKRIKAC 19

```

```

RESULT 10
US-09-114-637-2
; Sequence 2, Application US/09114637
; Patent No. 5945339
; GENERAL INFORMATION:
; APPLICANT: Kmiec, Eric
; APPLICANT: Holloman, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS TO PROMOTE
; TITLE OF INVENTION: HOMOLOGOUS RECOMBINATION IN EUKARYOTIC CELLS AND
ORGANISMS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/114,637
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/373,134
; FILING DATE: January 17, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Friebel, Thomas E.
; REGISTRATION NUMBER: 29,258
; REFERENCE/DOCKET NUMBER: 7991-007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 781 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-114-637-2

```

```

Query Match          69.4%; Score 34; DB 2; Length 781;
Best Local Similarity 55.6%; Pred. No. 1.3e+02;
Matches      5; Conservative      2; Mismatches      2; Indels      0; Gaps      0;

```

```

Qy      1 CLSSRLDAC 9
        |:| |:|
Db      11 CISKRIKAC 19

```

```

RESULT 11
US-08-530-010-33
; Sequence 33, Application US/08530010
; Patent No. 5689055
; GENERAL INFORMATION:
; APPLICANT: Meyerowitz, Elliott M.
; APPLICANT: Chang, Caren
; APPLICANT: Bleecker, Anthony B.
; TITLE OF INVENTION: PLANTS HAVING MODIFIED RESPONSE TO ETHYLENE
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Richard F. Trecartin
; STREET: 3400 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,010
; FILING DATE:

```

```

; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/086,555
; FILING DATE: 01-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-57515/RFT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-530-010-33

```

```

Query Match          67.3%; Score 33; DB 1; Length 44;
Best Local Similarity 55.6%; Pred. No. 10;
Matches      5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy      1 CLSSRLDAC 9
        || | :|:|
Db      18 CLESGMDSC 26

```

# RESULT 12

US-08-484-101B-33

; Sequence 33, Application US/08484101B

; Patent No. 5824868

; GENERAL INFORMATION:

; APPLICANT: California Institute of Technology

; TITLE OF INVENTION: PLANTS HAVING MODIFIED RESPONSE TO

; TITLE OF INVENTION: ETHYLENE

; NUMBER OF SEQUENCES: 50

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Richard F. Trecartin

; STREET: 3400 Embarcadero Center, Suite 3400

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94111

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/484,101B

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/US94/

; FILING DATE: 01-JUL-1994

```

; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/086,555
; FILING DATE: 01-JUL-1993
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-57515-2/RFT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-484-101B-33

```

```

Query Match          67.3%; Score 33; DB 2; Length 44;
Best Local Similarity 55.6%; Pred. No. 10;
Matches      5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy      1 CLSSRLDAC 9
        || | :|:|
Db      18 CLESGMDSC 26

```

```

RESULT 13
US-08-714-524D-33
; Sequence 33, Application US/08714524D
; Patent No. 6294716
; GENERAL INFORMATION:
; APPLICANT: Meyerowitz, Elliott M
; APPLICANT: Chang, Caren
; APPLICANT: Bleecker, Anthony B
; TITLE OF INVENTION: PLANTS HAVING MODIFIED RESPONSE TO ETHYLENE
; FILE REFERENCE: a-57515-4
; CURRENT APPLICATION NUMBER: US/08/714,524D
; CURRENT FILING DATE: 1996-09-16
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 44
; TYPE: PRT
; ORGANISM: Escherichia coli
US-08-714-524D-33

```

```

Query Match          67.3%; Score 33; DB 3; Length 44;
Best Local Similarity 55.6%; Pred. No. 10;
Matches      5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy      1 CLSSRLDAC 9
        || | :|:|
Db      18 CLESGMDSC 26

```



RESULT 14  
 US-08-969-644-20  
 ; Sequence 20, Application US/08969644  
 ; Patent No. 6096519  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ratti, Giulio  
 ; APPLICANT: Comanducci, Maurizio  
 ; APPLICANT: Tecce, Mario F.  
 ; APPLICANT: Giuliani, Marzia M.  
 ; TITLE OF INVENTION: pCTD PLASMID ISOLATED FROM CHLAMYDIA  
 ; TITLE OF INVENTION: TRACHOMATIS SEROTYPE D, ITS GENES AND PROTEINS ENCODED  
 BY  
 ; TITLE OF INVENTION: THEM; RECOMBINANT PLASMIDS FOR THE EXPRESSION OF SAID  
 ; TITLE OF INVENTION: GENES IN HETEROLOGOUS SYSTEMS, PREPARATION OF SAID  
 ; NUMBER OF SEQUENCES: 23  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH  
 ; STREET: 301 N. Washington Street  
 ; CITY: Falls Church  
 ; STATE: Virginia  
 ; COUNTRY: USA  
 ; ZIP: 22046-0747  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/969,644  
 ; FILING DATE: 13-NOV-1997  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/467,152  
 ; FILING DATE:  
 ; APPLICATION NUMBER: US/07/661,820  
 ; FILING DATE:  
 ; APPLICATION NUMBER: IT MI 91A000314  
 ; FILING DATE: 07-FEB-1991  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Svensson, Leonard R.  
 ; REGISTRATION NUMBER: 30,330  
 ; REFERENCE/DOCKET NUMBER: 1267-202P  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 703-241-1300  
 ; TELEFAX: 703-241-2848  
 ; TELEX: 248345  
 ; INFORMATION FOR SEQ ID NO: 20:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 309 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 US-08-969-644-20

Query Match

67.3%; Score 33; DB 3; Length 309;

Best Local Similarity 66.7%; Pred. No. 78;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9  
||| |  
Db 283 CLSSRQSV 291

RESULT 15

US-08-444-189-20

; Sequence 20, Application US/08444189

; Patent No. 6110705

; GENERAL INFORMATION:

; APPLICANT: Ratti, Giulio

; APPLICANT: Comanducci, Maurizio

; APPLICANT: Tecce, Mario F.

; APPLICANT: Giuliani, Marzia M.

; TITLE OF INVENTION: pCTD PLASMID ISOLATED FROM CHLAMYDIA

; TITLE OF INVENTION: TRACHOMATIS SEROTYPE D, ITS GENES AND PROTEINS ENCODED  
BY

; TITLE OF INVENTION: THEM; RECOMBINANT PLASMIDS FOR THE EXPRESSION OF SAID

; TITLE OF INVENTION: GENES IN HETEROLOGOUS SYSTEMS, PREPARATION OF SAID

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH

; STREET: 301 N. Washington Street

; CITY: Falls Church

; STATE: Virginia

; COUNTRY: USA

; ZIP: 22046-0747

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/444,189

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/180,528

; FILING DATE:

; APPLICATION NUMBER: US/07/991,512

; FILING DATE:

; APPLICATION NUMBER: US/07/661,820

; FILING DATE:

; APPLICATION NUMBER: IT MI 91A000314

; FILING DATE: 07-FEB-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Svensson, Leonard R.

; REGISTRATION NUMBER: 30,330

; REFERENCE/DOCKET NUMBER: 1267-202P

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-241-1300

; TELEFAX: 703-241-2848

; TELEX: 248345

; INFORMATION FOR SEQ ID NO: 20:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 309 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-444-189-20

Query Match 67.3%; Score 33; DB 3; Length 309;  
Best Local Similarity 66.7%; Pred. No. 78;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CLSSRLDAC 9  
||| |  
Db 283 CLSSRQSVC 291

Search completed: November 13, 2003, 09:54:56  
Job time : 11.6875 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:39:50 ; Search time 9.5 Seconds  
(without alignments)  
35.630 Million cell updates/sec

Title: US-09-228-866-4  
Perfect score: 48  
Sequence: 1 CVLRGGRC 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	48	100.0	8	1 US-08-526-710-4	Sequence 4, Appli
2	48	100.0	8	3 US-08-862-855-4	Sequence 4, Appli
3	48	100.0	8	3 US-09-226-985-4	Sequence 4, Appli
4	48	100.0	8	4 US-09-227-906-4	Sequence 4, Appli
5	38	79.2	34	4 US-09-636-399A-64	Sequence 64, Appl
6	38	79.2	35	4 US-09-636-399A-62	Sequence 62, Appl
7	38	79.2	35	4 US-09-636-399A-63	Sequence 63, Appl
8	38	79.2	36	4 US-09-636-399A-60	Sequence 60, Appl
9	38	79.2	36	4 US-09-636-399A-61	Sequence 61, Appl
10	38	79.2	37	4 US-09-636-399A-58	Sequence 58, Appl
11	38	79.2	37	4 US-09-636-399A-59	Sequence 59, Appl

12	38	79.2	38	4	US-09-636-399A-56	Sequence 56, Appl
13	38	79.2	38	4	US-09-636-399A-57	Sequence 57, Appl
14	38	79.2	39	4	US-09-636-399A-54	Sequence 54, Appl
15	38	79.2	39	4	US-09-636-399A-55	Sequence 55, Appl
16	38	79.2	40	4	US-09-636-399A-52	Sequence 52, Appl
17	38	79.2	40	4	US-09-636-399A-53	Sequence 53, Appl
18	38	79.2	41	4	US-09-636-399A-50	Sequence 50, Appl
19	38	79.2	41	4	US-09-636-399A-51	Sequence 51, Appl
20	38	79.2	42	4	US-09-636-399A-48	Sequence 48, Appl
21	38	79.2	42	4	US-09-636-399A-49	Sequence 49, Appl
22	38	79.2	43	4	US-09-636-399A-46	Sequence 46, Appl
23	38	79.2	43	4	US-09-636-399A-47	Sequence 47, Appl
24	38	79.2	44	4	US-09-636-399A-44	Sequence 44, Appl
25	38	79.2	44	4	US-09-636-399A-45	Sequence 45, Appl
26	38	79.2	45	4	US-09-868-659-2	Sequence 2, Appli
27	38	79.2	45	4	US-09-636-399A-42	Sequence 42, Appl
28	38	79.2	45	4	US-09-636-399A-43	Sequence 43, Appl
29	38	79.2	46	4	US-09-636-399A-40	Sequence 40, Appl
30	38	79.2	46	4	US-09-636-399A-41	Sequence 41, Appl
31	38	79.2	47	4	US-09-636-399A-38	Sequence 38, Appl
32	38	79.2	47	4	US-09-636-399A-39	Sequence 39, Appl
33	38	79.2	48	4	US-09-636-399A-36	Sequence 36, Appl
34	38	79.2	48	4	US-09-636-399A-37	Sequence 37, Appl
35	38	79.2	49	4	US-09-636-399A-35	Sequence 35, Appl
36	38	79.2	65	4	US-09-636-399A-2	Sequence 2, Appli
37	38	79.2	67	4	US-09-868-659-4	Sequence 4, Appli
38	38	79.2	67	4	US-09-636-399A-10	Sequence 10, Appl
39	36	75.0	307	1	US-08-164-614A-11	Sequence 11, Appl
40	36	75.0	307	2	US-08-456-489B-11	Sequence 11, Appl
41	36	75.0	340	4	US-09-252-991A-26708	Sequence 26708, A
42	36	75.0	522	1	US-08-164-614A-10	Sequence 10, Appl
43	36	75.0	522	2	US-08-456-489B-10	Sequence 10, Appl
44	36	75.0	536	1	US-08-164-614A-12	Sequence 12, Appl
45	36	75.0	536	2	US-08-456-489B-12	Sequence 12, Appl

#### ALIGNMENTS

##### RESULT 1

US-08-526-710-4

; Sequence 4, Application US/08526710

; Patent No. 5622699

; GENERAL INFORMATION:

; APPLICANT: Ruoslahti, Erkki

; APPLICANT: Pasqualini, Renata

; TITLE OF INVENTION: Method of Identifying Molecules That

; TITLE OF INVENTION: Home to a Selected Organ In Vivo

; NUMBER OF SEQUENCES: 44

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Campbell and Flores

; STREET: 4370 La Jolla Village Drive, Suite 700

; CITY: San Diego

; STATE: California

; COUNTRY: United States

; ZIP: 92122

; COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/526,710
; FILING DATE: 11-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-526-710-4

```

```

Query Match          100.0%; Score 48; DB 1; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      8; Conservative    0; Mismatches    0; Indels      0; Gaps      0;

```

```

Qy      1 CVLRGGRC 8
        |||||
Db      1 CVLRGGRC 8

```

## RESULT 2

US-08-862-855-4

```

; Sequence 4, Application US/08862855
; Patent No. 6068829
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/862,855
; FILING DATE:

```

```

; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2621
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-862-855-4

```

```

Query Match          100.0%; Score 48; DB 3; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 CVLRGGRC 8
        |||||
Db      1 CVLRGGRC 8

```

```

RESULT 3
US-09-226-985-4
; Sequence 4, Application US/09226985
; Patent No. 6296832
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Molecules That Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/226,985
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,855
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3423
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-226-985-4

```

```

Query Match          100.0%; Score 48; DB 3; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      8; Conservative    0; Mismatches    0; Indels      0; Gaps      0;

```

```

Qy      1 CVLRGGRC 8
        |||||
Db      1 CVLRGGRC 8

```

#### RESULT 4

US-09-227-906-4

```

; Sequence 4, Application US/09227906
; Patent No. 6306365
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Hometo a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/227,906
; FILING DATE:

```



```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,855
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3424
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-227-906-4

```

```

Query Match          100.0%; Score 48; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      1 CVLRGGRC 8
        |||||
Db      1 CVLRGGRC 8

```

```

RESULT 5
US-09-636-399A-64
; Sequence 64, Application US/09636399A
; Patent No. 6576755
; GENERAL INFORMATION:
; APPLICANT: Adler, David A.
; APPLICANT: Holloway, James L.
; APPLICANT: Baindur, Nand
; APPLICANT: Beigel-Orme, Stephanie
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: NOVEL BETA-DEFENSINS
; FILE REFERENCE: 97-44C2
; CURRENT APPLICATION NUMBER: US/09/636,399A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/058,335
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/064,294
; PRIOR FILING DATE: 1997-11-05
; PRIOR APPLICATION NUMBER: 09/150,786
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 09/636,399
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 72

```

```
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 64
;   LENGTH: 34
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Defensin polypeptide
;   NAME/KEY: VARIANT
;   LOCATION: (31)...(31)
;   OTHER INFORMATION: Xaa is Ile, Leu, Val, Phe, or Met
US-09-636-399A-64
```

```
Query Match          79.2%;   Score 38;   DB 4;   Length 34;
Best Local Similarity 75.0%;   Pred. No. 5.3;
Matches      6;   Conservative      1;   Mismatches      1;   Indels      0;   Gaps      0;
```

```
Qy      1 CVLRGGRC 8
        | : |||||
Db      1 CRVRGGRC 8
```

# RESULT 6

```
US-09-636-399A-62
; Sequence 62, Application US/09636399A
; Patent No. 6576755
; GENERAL INFORMATION:
; APPLICANT: Adler, David A.
; APPLICANT: Holloway, James L.
; APPLICANT: Baindur, Nand
; APPLICANT: Beigel-Orme, Stephanie
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: NOVEL BETA-DEFENSINS
; FILE REFERENCE: 97-44C2
; CURRENT APPLICATION NUMBER: US/09/636,399A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/058,335
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/064,294
; PRIOR FILING DATE: 1997-11-05
; PRIOR APPLICATION NUMBER: 09/150,786
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 09/636,399
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 62
;   LENGTH: 35
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Defensin polypeptide
;   NAME/KEY: VARIANT
;   LOCATION: (32)...(32)
;   OTHER INFORMATION: Xaa is Phe, Val, Ile, Leu, or Met
US-09-636-399A-62
```

```
Query Match          79.2%;   Score 38;   DB 4;   Length 35;
```

Best Local Similarity 75.0%; Pred. No. 5.4;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
| :|||||  
Db 2 CRVRGGRC 9

RESULT 7

US-09-636-399A-63  
; Sequence 63, Application US/09636399A  
; Patent No. 6576755  
; GENERAL INFORMATION:  
; APPLICANT: Adler, David A.  
; APPLICANT: Holloway, James L.  
; APPLICANT: Baindur, Nand  
; APPLICANT: Beigel-Orme, Stephanie  
; APPLICANT: Sheppard, Paul O.  
; TITLE OF INVENTION: NOVEL BETA-DEFENSINS  
; FILE REFERENCE: 97-44C2  
; CURRENT APPLICATION NUMBER: US/09/636,399A  
; CURRENT FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/058,335  
; PRIOR FILING DATE: 1997-10-09  
; PRIOR APPLICATION NUMBER: 60/064,294  
; PRIOR FILING DATE: 1997-11-05  
; PRIOR APPLICATION NUMBER: 09/150,786  
; PRIOR FILING DATE: 1998-09-10  
; PRIOR APPLICATION NUMBER: 09/636,399  
; PRIOR FILING DATE: 2000-08-10  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 63  
; LENGTH: 35  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Defensin polypeptide  
; NAME/KEY: VARIANT  
; LOCATION: (31)...(31)  
; OTHER INFORMATION: Xaa is Ile, Leu, Phe, Val, or Met  
US-09-636-399A-63

Query Match 79.2%; Score 38; DB 4; Length 35;  
Best Local Similarity 75.0%; Pred. No. 5.4;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
| :|||||  
Db 1 CRVRGGRC 8

RESULT 8

US-09-636-399A-60  
; Sequence 60, Application US/09636399A  
; Patent No. 6576755  
; GENERAL INFORMATION:

```

; APPLICANT: Adler, David A.
; APPLICANT: Holloway, James L.
; APPLICANT: Baindur, Nand
; APPLICANT: Beigel-Orme, Stephanie
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: NOVEL BETA-DEFENSINS
; FILE REFERENCE: 97-44C2
; CURRENT APPLICATION NUMBER: US/09/636,399A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/058,335
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/064,294
; PRIOR FILING DATE: 1997-11-05
; PRIOR APPLICATION NUMBER: 09/150,786
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 09/636,399
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 60
;   LENGTH: 36
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Defensin polypeptide
;   NAME/KEY: VARIANT
;   LOCATION: (33)...(33)
;   OTHER INFORMATION: Xaa is Ile, Leu, Val, Phe, or Met
US-09-636-399A-60

```

```

Query Match          79.2%;   Score 38;   DB 4;   Length 36;
Best Local Similarity 75.0%;   Pred. No. 5.5;
Matches      6;   Conservative    1;   Mismatches    1;   Indels      0;   Gaps      0;

```

```

Qy      1 CVLRGGRC 8
        | :|||||
Db      3 CRVRGGRC 10

```

# RESULT 9

```

US-09-636-399A-61
; Sequence 61, Application US/09636399A
; Patent No. 6576755
; GENERAL INFORMATION:
; APPLICANT: Adler, David A.
; APPLICANT: Holloway, James L.
; APPLICANT: Baindur, Nand
; APPLICANT: Beigel-Orme, Stephanie
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: NOVEL BETA-DEFENSINS
; FILE REFERENCE: 97-44C2
; CURRENT APPLICATION NUMBER: US/09/636,399A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/058,335
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/064,294
; PRIOR FILING DATE: 1997-11-05

```

```

; PRIOR APPLICATION NUMBER: 09/150,786
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 09/636,399
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 61
;   LENGTH: 36
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Defensin polypeptide
;   NAME/KEY: VARIANT
;   LOCATION: (32)...(32)
;   OTHER INFORMATION: Xaa is Leu, Ile, Val, Met, or Phe
US-09-636-399A-61

```

```

Query Match          79.2%;   Score 38;   DB 4;   Length 36;
Best Local Similarity 75.0%;   Pred. No. 5.5;
Matches      6;   Conservative    1;   Mismatches    1;   Indels      0;   Gaps      0;

```

```

Qy      1 CVLRGGRC 8
        | :|||||
Db      2 CRVRGGRC 9

```

```

RESULT 10
US-09-636-399A-58
; Sequence 58, Application US/09636399A
; Patent No. 6576755
; GENERAL INFORMATION:
; APPLICANT: Adler, David A.
; APPLICANT: Holloway, James L.
; APPLICANT: Baindur, Nand
; APPLICANT: Beigel-Orme, Stephanie
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: NOVEL BETA-DEFENSINS
; FILE REFERENCE: 97-44C2
; CURRENT APPLICATION NUMBER: US/09/636,399A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/058,335
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/064,294
; PRIOR FILING DATE: 1997-11-05
; PRIOR APPLICATION NUMBER: 09/150,786
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 09/636,399
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 58
;   LENGTH: 37
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Defensin polypeptide
;   NAME/KEY: VARIANT

```

; LOCATION: (34)...(34)  
; OTHER INFORMATION: Xaa is Ile, Leu, Val, Phe, or Met  
US-09-636-399A-58

Query Match 79.2%; Score 38; DB 4; Length 37;  
Best Local Similarity 75.0%; Pred. No. 5.7;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
| :|||||  
Db 4 CRVRGGRC 11

RESULT 11

US-09-636-399A-59  
; Sequence 59, Application US/09636399A  
; Patent No. 6576755  
; GENERAL INFORMATION:  
; APPLICANT: Adler, David A.  
; APPLICANT: Holloway, James L.  
; APPLICANT: Baindur, Nand  
; APPLICANT: Beigel-Orme, Stephanie  
; APPLICANT: Sheppard, Paul O.  
; TITLE OF INVENTION: NOVEL BETA-DEFENSINS  
; FILE REFERENCE: 97-44C2  
; CURRENT APPLICATION NUMBER: US/09/636,399A  
; CURRENT FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/058,335  
; PRIOR FILING DATE: 1997-10-09  
; PRIOR APPLICATION NUMBER: 60/064,294  
; PRIOR FILING DATE: 1997-11-05  
; PRIOR APPLICATION NUMBER: 09/150,786  
; PRIOR FILING DATE: 1998-09-10  
; PRIOR APPLICATION NUMBER: 09/636,399  
; PRIOR FILING DATE: 2000-08-10  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 59  
; LENGTH: 37  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Defensin polypeptide  
; NAME/KEY: VARIANT  
; LOCATION: (33)...(33)  
; OTHER INFORMATION: Xaa is Ile, Leu, Met, Phe, or Val  
US-09-636-399A-59

Query Match 79.2%; Score 38; DB 4; Length 37;  
Best Local Similarity 75.0%; Pred. No. 5.7;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
| :|||||  
Db 3 CRVRGGRC 10

RESULT 12

US-09-636-399A-56

```
; Sequence 56, Application US/09636399A
; Patent No. 6576755
; GENERAL INFORMATION:
; APPLICANT: Adler, David A.
; APPLICANT: Holloway, James L.
; APPLICANT: Baindur, Nand
; APPLICANT: Beigel-Orme, Stephanie
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: NOVEL BETA-DEFENSINS
; FILE REFERENCE: 97-44C2
; CURRENT APPLICATION NUMBER: US/09/636,399A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/058,335
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/064,294
; PRIOR FILING DATE: 1997-11-05
; PRIOR APPLICATION NUMBER: 09/150,786
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 09/636,399
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 56
; LENGTH: 38
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Defensin polypeptide
; NAME/KEY: VARIANT
; LOCATION: (35)...(35)
; OTHER INFORMATION: Xaa is Ile, Leu, Val, Phe, or Met
US-09-636-399A-56
```

```
Query Match          79.2%; Score 38; DB 4; Length 38;
Best Local Similarity 75.0%; Pred. No. 5.8;
Matches      6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy      1 CVLRGGRC 8
        | :|||||
Db      5 CRVRGGRC 12
```

RESULT 13

US-09-636-399A-57

```
; Sequence 57, Application US/09636399A
; Patent No. 6576755
; GENERAL INFORMATION:
; APPLICANT: Adler, David A.
; APPLICANT: Holloway, James L.
; APPLICANT: Baindur, Nand
; APPLICANT: Beigel-Orme, Stephanie
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: NOVEL BETA-DEFENSINS
; FILE REFERENCE: 97-44C2
; CURRENT APPLICATION NUMBER: US/09/636,399A
```

```

; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/058,335
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/064,294
; PRIOR FILING DATE: 1997-11-05
; PRIOR APPLICATION NUMBER: 09/150,786
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 09/636,399
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 57
; LENGTH: 38
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Defensin polypeptide
; NAME/KEY: VARIANT
; LOCATION: (34)...(34)
; OTHER INFORMATION: Xaa is Ile, Leu, Val, Phe, or Met
US-09-636-399A-57

```

```

Query Match          79.2%; Score 38; DB 4; Length 38;
Best Local Similarity 75.0%; Pred. No. 5.8;
Matches      6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 CVLRGGRC 8
        | :|||||
Db      4 CRVRGGRC 11

```

```

RESULT 14
US-09-636-399A-54
; Sequence 54, Application US/09636399A
; Patent No. 6576755
; GENERAL INFORMATION:
; APPLICANT: Adler, David A.
; APPLICANT: Holloway, James L.
; APPLICANT: Baindur, Nand
; APPLICANT: Beigel-Orme, Stephanie
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: NOVEL BETA-DEFENSINS
; FILE REFERENCE: 97-44C2
; CURRENT APPLICATION NUMBER: US/09/636,399A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/058,335
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/064,294
; PRIOR FILING DATE: 1997-11-05
; PRIOR APPLICATION NUMBER: 09/150,786
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 09/636,399
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 54
; LENGTH: 39

```



```
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Defensin polypeptide
; NAME/KEY: VARIANT
; LOCATION: (36)...(36)
; OTHER INFORMATION: Xaa is Leu, Ile, Met, Phe, or Val
US-09-636-399A-54
```

```
Query Match          79.2%;  Score 38;  DB 4;  Length 39;
Best Local Similarity 75.0%;  Pred. No. 6;
Matches      6;  Conservative      1;  Mismatches      1;  Indels      0;  Gaps      0;
```

```
Qy      1 CVLRGGRC 8
        | :|||||
Db      6 CRVRGGRC 13
```

# RESULT 15

```
US-09-636-399A-55
; Sequence 55, Application US/09636399A
; Patent No. 6576755
; GENERAL INFORMATION:
; APPLICANT: Adler, David A.
; APPLICANT: Holloway, James L.
; APPLICANT: Baindur, Nand
; APPLICANT: Beigel-Orme, Stephanie
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: NOVEL BETA-DEFENSINS
; FILE REFERENCE: 97-44C2
; CURRENT APPLICATION NUMBER: US/09/636,399A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/058,335
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/064,294
; PRIOR FILING DATE: 1997-11-05
; PRIOR APPLICATION NUMBER: 09/150,786
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 09/636,399
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 55
; LENGTH: 39
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Defensin polypeptide
; NAME/KEY: VARIANT
; LOCATION: (35)...(35)
; OTHER INFORMATION: Xaa is Leu, Val, Ile, Met, or Phe
US-09-636-399A-55
```

```
Query Match          79.2%;  Score 38;  DB 4;  Length 39;
Best Local Similarity 75.0%;  Pred. No. 6;
Matches      6;  Conservative      1;  Mismatches      1;  Indels      0;  Gaps      0;
```

Qy            1 CVLRGGRC 8  
              | :|||||  
Db            5 CRVRGGRC 12

Search completed: November 13, 2003, 09:54:56  
Job time : 9.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:31:40 ; Search time 26.9167 Seconds  
(without alignments)  
47.176 Million cell updates/sec

Title: US-09-228-866-4  
Perfect score: 48  
Sequence: 1 CVLRGGRC 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_19Jun03:\*  
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:\*  
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:\*  
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:\*  
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:\*  
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:\*  
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:\*  
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:\*  
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:\*  
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:\*  
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:\*  
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:\*  
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:\*  
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:\*  
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:\*  
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:\*  
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:\*  
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:\*  
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:\*  
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:\*  
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:\*  
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:\*  
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:\*  
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:\*  
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	48	100.0	8	18	AAW13414	Brain homing pepti
2	48	100.0	8	21	AAB07390	Brain homing pepti
3	48	100.0	8	22	AAE11796	Phage peptide #4 t
4	48	100.0	8	23	AAU10707	Brain homing pepti
5	42	87.5	537	22	ABG25335	Novel human diagno
6	41	85.4	36	24	ABU61386	Human A domain fro
7	41	85.4	4561	22	ABG30203	Novel human diagno
8	41	85.4	9222	22	ABG21064	Novel human diagno
9	40	83.3	175	21	AAB42272	Human ORFX ORF2036
10	39	81.2	105	22	AAU53302	Propionibacterium
11	39	81.2	2743	23	ABB81598	Human laminin alph
12	39	81.2	3695	23	ABB81588	Human laminin alph
13	39	81.2	3696	23	AAE17310	Human laminin alph
14	39	81.2	3705	23	AAE17309	Human laminin alph
15	38	79.2	8	23	AAO17769	Human beta-defensi
16	38	79.2	17	23	AAO17771	Human beta-defensi
17	38	79.2	17	23	AAO17780	Human beta-defensi
18	38	79.2	22	23	AAO17772	Human beta-defensi
19	38	79.2	31	23	AAO17765	Human beta-defensi
20	38	79.2	31	23	AAM49572	Human beta-defensi
21	38	79.2	31	23	AAM49576	Human beta-defensi
22	38	79.2	32	21	AAB10621	Human SAP-3 N-term
23	38	79.2	40	23	AAO17766	Human beta-defensi
24	38	79.2	41	23	AAU09708	Human beta-defensi
25	38	79.2	45	21	AAB10600	Human SAP-3 mature
26	38	79.2	45	23	AAO17767	Human beta-defensi
27	38	79.2	45	23	AAU09709	Human beta-defensi
28	38	79.2	65	20	AAY07243	Beta-defensin fami
29	38	79.2	67	20	AAY07244	Beta-defensin fami
30	38	79.2	67	21	AAB10602	Human SAP-3 pre-pr
31	38	79.2	67	23	AAO17768	Human beta-defensi
32	38	79.2	67	23	AAU91016	Transplant media a
33	38	79.2	67	23	AAU91036	Transplant media a
34	38	79.2	67	23	AAU09707	Human beta-defensi
35	38	79.2	19938	24	ABP76681	Streptomyces virid
36	37	77.1	226	22	AAU53350	Propionibacterium
37	37	77.1	376	21	AAY75053	Neisseria gonorrhe
38	37	77.1	387	24	ABP77315	N. gonorrhoeae ami
39	37	77.1	478	22	AAU58991	Propionibacterium
40	36	75.0	41	22	AAB86262	Murine beta-defens
41	36	75.0	63	22	AAE02126	Mouse beta defensi
42	36	75.0	65	21	AAB23178	Phytolacca america
43	36	75.0	110	22	ABB11336	Human PRGE-30 homo
44	36	75.0	128	19	AAW64060	Human IL-9 recepto
45	36	75.0	150	19	AAW64061	Human IL-9 recepto

# ALIGNMENTS

RESULT 1

AAW13414

ID AAW13414 standard; Peptide; 8 AA.

XX

AC AAW13414;

XX

DT 15-JAN-1998 (first entry)

XX

DE Brain homing peptide.

XX

KW Brain homing peptide; in vivo panning; screening; phage display;

KW drug delivery.

XX

OS Synthetic.

XX

PN WO9710507-A1.

XX

PD 20-MAR-1997.

XX

PF 10-SEP-1996; 96WO-US14600.

XX

PR 11-SEP-1995; 95US-0526710.

PR 11-SEP-1995; 95US-0526708.

XX

PA (LJOL-) LA JOLLA CANCER RES FOUND.

XX

PI Pasqualini R, Ruoslahti E;

XX

DR WPI; 1997-202359/18.

XX

PT Obtaining compound that homes to selected organ or tissue - by in

PT vivo panning method, specifically to identify brain, kidney,

PT angiogenic vasculature or tumour tissue homing peptide(s)

XX

PS Claim 13; Page 67; 75pp; English.

XX

CC This synthetic peptide is a claimed example of a brain-homing  
 CC peptide that was identified using a novel method for obtaining  
 CC molecules that home to a selected organ or tissue. This in vivo  
 CC panning method typically involves administering a phage display  
 CC library to a subject, and identifying expressed peptides which  
 CC home to the desired organ or tissue, e.g. brain, kidney, angiogenic  
 CC vascular tissue or tumour tissue. The isolated peptides (see  
 CC AAW13412-52, AAW11181-86) can be used to target e.g. drugs, toxins or  
 CC labels to the selected organ/tissue (claimed) or to identify and/or  
 CC isolate target molecules (claimed). The peptides can be directly  
 CC identified in vivo, as compared to prior art in vitro screening  
 CC methods, which require further examination to see if they maintain  
 CC specificity in vivo.

XX

SQ Sequence 8 AA;

Query Match 100.0%; Score 48; DB 18; Length 8;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8

Db                    |||||  
1 CVLRGGRC 8

RESULT 2

AAB07390

ID    AAB07390 standard; peptide; 8 AA.

XX

AC    AAB07390;

XX

DT    17-OCT-2000    (first entry)

XX

DE    Brain homing peptide # 4.

XX

KW    Brain; homing peptide; organ targeting; tissue targeting; mouse; cyclic.

XX

OS    Mus sp.

XX

FH    Key                    Location/Qualifiers

FT    Disulfide-bond    1..8

FT                    /note= "Can optionally form a cyclic peptide"

XX

PN    US6068829-A.

XX

PD    30-MAY-2000.

XX

PF    23-JUN-1997;    97US-0862855.

XX

PR    11-SEP-1995;    95US-0526710.

PR    10-MAR-1997;    97US-0813273.

XX

PA    (BURN-) BURNHAM INST.

XX

PI    Pasqualini R,    Ruoslahti E;

XX

DR    WPI; 2000-410850/35.

XX

PT    Identifying and recovering organ homing molecules or peptides by in

PT    vivo panning comprises administering a library of diverse peptides

PT    linked to a tag which facilitates recovery of these peptides -

XX

PS    Example 2; Column 17; 20pp; English.

XX

CC    The present sequence is a mouse brain homing peptide. This sequence was

CC    identified by using in vivo panning to screen a library of potential

CC    organ homing molecules. The present sequence can be used to direct a

CC    moiety to a the brain tissue, by linking the moiety to the present

CC    sequence. Examples of potential moieties are drugs, toxins or a

CC    detectable label. The present sequence contains a VRL amino acid motif.

XX

SQ    Sequence    8 AA;

Query Match                    100.0%;    Score 48;    DB 21;    Length 8;

Best Local Similarity    100.0%;    Pred. No. 9.3e+05;

Matches    8;    Conservative    0;    Mismatches    0;    Indels    0;    Gaps    0;

Qy                    1 CVLRGGRC 8

Db                    |||||  
                      1 CVLRGGRC 8

RESULT 3

AAE11796

ID    AAE11796 standard; peptide; 8 AA.

XX

AC    AAE11796;

XX

DT    18-DEC-2001    (first entry)

XX

DE    Phage peptide #4 targetted to brain.

XX

KW    Enriched library fraction; brain; kidney; tumour; panning; diagnostic;

KW    molecular medicine; drug delivery; peptidomimetic; pharmaceutical.

XX

OS    Bacteriophage.

XX

FH    Key                    Location/Qualifiers

FT    Domain                2..4

FT                         /label= VLR\_motif

XX

PN    US6296832-B1.

XX

PD    02-OCT-2001.

XX

PF    08-JAN-1999;    99US-0226985.

XX

PR    23-JUN-1997;    97US-0862855.

PR    11-SEP-1995;    95US-0526710.

PR    10-MAR-1997;    97US-0813273.

XX

PA    (BURN-) BURNHAM INST.

XX

PI    Ruoslahti E,    Pasqualini R;

XX

DR    WPI; 2001-610691/70.

XX

PT    Enriched library fraction comprising molecules recovered by in vivo

PT    panning that selectively home to a selected organ or tissue useful for

PT    treating disease or in diagnostic methods    -

XX

PS    Example 2; Column 17; 21pp; English.

XX

CC    The invention relates to an enriched library fraction containing  
CC    molecules that selectively home to a selected organ or tissue such as  
CC    brain, kidney or tumour recovered by in vivo panning. The invention  
CC    generally relates to the field of molecular medicine, drug delivery and  
CC    to a method of invivo panning for identifying a molecule that homes to a  
CC    specific organ. The molecules, e.g., peptides, peptidomimetics, proteins  
CC    and fragments of proteins contained in an enriched library fraction may  
CC    be administered to a subject as part of a pharmaceutical composition to  
CC    treat disease or in diagnostic methods. The present sequence is a  
CC    peptide from bacteriophage targetted to brain.

XX

SQ    Sequence    8 AA;

Query Match 100.0%; Score 48; DB 22; Length 8;  
Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
|||  
Db 1 CVLRGGRC 8

RESULT 4

AAU10707

ID AAU10707 standard; peptide; 8 AA.

XX

AC AAU10707;

XX

DT 12-MAR-2002 (first entry)

XX

DE Brain homing peptide #4 useful for delivery of target molecules.

XX

KW Organ targeting; tissue targeting; cancer; tumour homing molecule;  
KW delivery of target molecule; brain homing peptide.

XX

OS Synthetic.

XX

PN US6306365-B1.

XX

PD 23-OCT-2001.

XX

PF 08-JAN-1999; 99US-0227906.

XX

PR 23-JUN-1997; 97US-0862855.

PR 11-SEP-1995; 95US-0526710.

PR 10-MAR-1997; 97US-0813273.

XX

PA (BURN-) BURNHAM INST.

XX

PI Ruoslahti E, Pasqualini R;

XX

DR WPI; 2002-040196/05.

XX

PT Recovering molecules that home to an organ or tissue, useful for  
PT identifying molecules that home to a specific organ or tissue, e.g.  
PT identifying a tumour homing molecule to identify the presence of cancer,  
PT by in vivo panning of a library -

XX

PS Example 2; Column 17; 21pp; English.

XX

CC The present invention relates to a method of recovering molecules that  
CC home to a selected organ or tissue. The method comprises administering  
CC to the subject the library of diverse molecules, collecting a sample of  
CC the selected organ or tissue (e.g. brain or kidney), and recovering from  
CC the sample several molecules that home to the selected organ or tissue.  
CC The method is useful for identifying molecules, particularly useful for  
CC screening large number of molecules (e.g. peptides), that home to a  
CC specific organ. The identified molecule is useful for e.g. raising an  
CC antibody specific for a target molecule, targeting a desired moiety



CC (e.g. drug, toxin or detectable label) to the selected organ.  
CC Specifically, the method is useful for identifying the presence of cancer  
CC in a subject by linking an appropriate moiety to a tumour homing  
CC molecule. The present method provides a direct means for identifying  
CC molecules that specifically home to a selected organ and, therefore  
CC provides a significant advantage over previous methods, which require  
CC that a molecule identified using an in vitro screening method  
CC subsequently be examined to determine if it maintains its specificity in  
CC vivo. AAU10704-AAU10723 represent brain homing peptides described in  
CC the present invention.

XX

SQ Sequence 8 AA;

Query Match 100.0%; Score 48; DB 23; Length 8;  
Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
|||  
Db 1 CVLRGGRC 8

#### RESULT 5

ABG25335

ID ABG25335 standard; Protein; 537 AA.

XX

AC ABG25335;

XX

DT 18-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #25326.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US08631.

XX

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;

XX

DR WPI; 2001-639362/73.

DR N-PSDB; AAS89522.

XX

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

XX  
PS Claim 20; SEQ ID No 55694; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 537 AA;

Query Match 87.5%; Score 42; DB 22; Length 537;  
Best Local Similarity 87.5%; Pred. No. 55;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVLRGGRC 8  
||| |||  
Db 348 CVLSGGRC 355

RESULT 6  
ABU61386  
ID ABU61386 standard; Peptide; 36 AA.  
XX  
AC ABU61386;  
XX  
DT 08-MAY-2003 (first entry)  
XX  
DE Human A domain from cDNA 075851 #8.  
XX  
KW LDL-receptor class A domain; A domain; human; domain multimer;  
KW multimer library; immuno-domain library.  
XX  
OS Homo sapiens.  
XX  
PN WO200288171-A2.  
XX  
PD 07-NOV-2002.  
XX  
PF 26-APR-2002; 2002WO-US13257.  
XX

PR 26-APR-2001; 2001US-286823P.  
PR 19-NOV-2001; 2001US-337209P.  
PR 26-NOV-2001; 2001US-333359P.  
PR 18-APR-2002; 2002US-374107P.

XX

PA (MAXY-) MAXYGEN INC.

XX

PI Kolkman JA, Stemmer WPC;

XX

DR WPI; 2003-111869/10.

XX

PT Identifying a multimer that binds to a target molecule, comprises  
PT identifying at least one monomer domain that bind to at least one  
PT target molecule and linking the identified monomer domains to form a  
PT library of multimers -

XX

PS Disclosure; Figure 10; 98pp; English.

XX

CC The invention relates to identifying a multimer that binds to a target  
CC molecule, comprising identifying at least one monomer domain that binds  
CC to at least one target molecule, linking the identified monomer domains  
CC to form a library of multimers, each multimer comprising at least two  
CC monomer domains, and screening the library of multimers for the ability  
CC to bind to the first target molecule. Also included are: (1) a library of  
CC multimers formed by the method above (where each multimer comprises at  
CC least two monomer domains connected by a linker, and each monomer domain  
CC exhibits a binding specificity for a target molecule); (2) a polypeptide  
CC comprising: (a) the multimer selected from the novel method; or (b) at  
CC least two monomer domains separated by a heterologous linker, where each  
CC monomer domain specifically binds to a target molecule; (3) a  
CC polynucleotide encoding the multimer selected from the novel method; and  
CC (4) identifying hetero-immuno multimers that bind to a target molecule,  
CC comprising: (a) providing a library of immuno-domains; (b) screening the  
CC library of immuno-domains for affinity to a first target molecule;  
CC (c) providing a library of monomer domains; (d) screening the library of  
CC monomer domains for affinity to a first target molecule; (e) identifying  
CC at least one immuno-domain that binds to at least one target molecule;  
CC (f) identifying at least one monomer domain that binds to at least one  
CC target molecule; (g) linking the identified immuno-domain with the  
CC identified monomer domains to form a library of multimers, each multimer  
CC comprising at least two domains; (h) screening the library of multimers  
CC for the ability to bind to the first target molecule; and (i) identifying  
CC a multimer that binds to the first target molecule. The methods are  
CC useful for identifying multimers that bind to target molecules. The  
CC methods can also be used for selecting and optimising properties of  
CC discrete monomer domains and/or immuno-domains to create multimers. The  
CC multimers are useful for identifying the multimers with improved  
CC phenotype such as improved avidity or affinity or altered specificity for  
CC the target molecule. The polynucleotide, polypeptide and/or multimer are  
CC useful for preventing or treating a disease or disorder in a subject. The  
CC present sequence is a human LDL (low density lipoprotein) class A domain  
CC or simply an A domain used to design a library of A domain multimers of  
CC the invention.

XX

SQ Sequence 36 AA;

Query Match

85.4%; Score 41; DB 24; Length 36;

Best Local Similarity 87.5%; Pred. No. 7.9;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
      |||||||  
Db 14 CVLRGGPC 21

RESULT 7

ABG30203

ID ABG30203 standard; Protein; 4561 AA.

XX

AC ABG30203;

XX

DT 18-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #30194.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US08631.

XX

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;

XX

DR WPI; 2001-639362/73.

DR N-PSDB; AAS94390.

XX

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

XX

PS Claim 20; SEQ ID No 60562; 103pp; English.

XX

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 4561 AA;

Query Match 85.4%; Score 41; DB 22; Length 4561;  
Best Local Similarity 87.5%; Pred. No. 5e+02;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
||| ||| |  
Db 1254 CVLRGGPC 1261

#### RESULT 8

ABG21064

ID ABG21064 standard; Protein; 9222 AA.

XX

AC ABG21064;

XX

DT 18-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #21055.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US08631.

XX

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;

XX

DR WPI; 2001-639362/73.

DR N-PSDB; AAS85251.

XX

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

XX  
 PS Claim 20; SEQ ID No 51423; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 9222 AA;

Query Match 85.4%; Score 41; DB 22; Length 9222;  
 Best Local Similarity 87.5%; Pred. No. 9.2e+02;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
 |||||  
 Db 1978 CVLRGGPC 1985

# RESULT 9

AAB42272

ID AAB42272 standard; Protein; 175 AA.

XX

AC AAB42272;

XX

DT 08-FEB-2001 (first entry)

XX

DE Human ORFX ORF2036 polypeptide sequence SEQ ID NO:4072.

XX

KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KW vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;

KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200058473-A2.  
 XX  
 PD 05-OCT-2000.  
 XX  
 PF 31-MAR-2000; 2000WO-US08621.  
 XX  
 PR 31-MAR-1999; 99US-0127607.  
 PR 02-APR-1999; 99US-0127636.  
 PR 05-APR-1999; 99US-0127728.  
 PR 30-MAR-2000; 2000US-0540763.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Shimkets RA, Leach M;  
 XX  
 DR WPI; 2000-602362/57.  
 DR N-PSDB; AAC76481.  
 XX  
 PT Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 XX  
 PS Claim 11; Page 3261; 5507pp; English.  
 XX  
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;  
 CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
 CC antithyroid; and antianaemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.  
 XX  
 SQ Sequence 175 AA;

Query Match 83.3%; Score 40; DB 21; Length 175;  
 Best Local Similarity 75.0%; Pred. No. 45;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy            1 CVLRGGRC 8  
             |: |||||  
Db            10 CLARGGRC 17

RESULT 10

AAU53302

ID    AAU53302 standard; Protein; 105 AA.

XX

AC    AAU53302;

XX

DT    27-FEB-2002    (first entry)

XX

DE    Propionibacterium acnes immunogenic protein #14198.

XX

KW    SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
KW    uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
KW    inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
KW    dermatological; osteopathic; neuroprotectant.

XX

OS    Propionibacterium acnes.

XX

PN    WO200181581-A2.

XX

PD    01-NOV-2001.

XX

PF    20-APR-2001; 2001WO-US12865.

XX

PR    21-APR-2000; 2000US-199047P.

PR    02-JUN-2000; 2000US-208841P.

PR    07-JUL-2000; 2000US-216747P.

XX

PA    (CORI-) CORIXA CORP.

XX

PI    Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

PI    L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX

DR    WPI; 2001-616774/71.

DR    N-PSDB; AAS59559.

XX

PT    Propionibacterium acnes polypeptides and nucleic acids useful for  
PT    vaccinating against and diagnosing infections, especially useful for  
PT    treating acne vulgaris -

XX

PS    Example 1; SEQ ID No 14497; 1069pp; English.

XX

CC    Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC    polypeptides. The proteins and their associated DNA sequences are used in  
CC    the treatment, prevention and diagnosis of medical conditions caused by  
CC    P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC    pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
CC    P. acnes is also involved in infections of bone, joints and the central  
CC    nervous system, however it is particularly involved in the inflammatory  
CC    lesions associated with acne vulgaris. A method for detecting the  
CC    presence or absence of P. acnes in a patient comprises contacting a  
CC    sample with a binding agent that binds to the proteins of the invention



CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 105 AA;

Query Match 81.2%; Score 39; DB 22; Length 105;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VLRGGRC 8  
|||  
Db 63 VLRGGRC 69

RESULT 11

ABB81598

ID ABB81598 standard; Protein; 2743 AA.

XX

AC ABB81598;

XX

DT 19-SEP-2002 (first entry)

XX

DE Human laminin alpha 5 2743 N-terminal amino acid sequence SEQ ID NO:36.

XX

KW Laminin alpha 5; laminin 10; vulnerary; cell growth; differentiation;  
KW tissue repair development; laminin; healing; vascular tissue;  
KW re-endothelialisation; vascular injury; cell attachment; cell stasis;  
KW proliferation; migration.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Peptide 1..35

FT /label= signal

FT Protein 36..2743

FT /label= laminin\_alpha\_5

XX

PN WO200250111-A2.

XX

PD 27-JUN-2002.

XX

PF 21-DEC-2001; 2001WO-US51035.

XX

PR 21-DEC-2000; 2000US-257449P.

PR 28-MAR-2001; 2001US-279282P.

PR 13-NOV-2001; 2001US-0279282.

XX

PA (BIOS-) BIOSTRATUM INC.

XX

PI Tryggvason K, Doi M, Thyboll J;  
 XX  
 DR WPI; 2002-557650/59.  
 DR N-PSDB; ABQ72930.  
 XX  
 PT New human laminin-10 proteins, useful for accelerating the healing of  
 PT vascular tissue, improving the biocompatibility of grafts, or for  
 PT promoting re-endothelialization at the site of vascular injuries -  
 XX  
 PS Disclosure; Page 223-231; 231pp; English.  
 XX  
 CC The present invention describes human laminin alpha 5. Also described  
 CC is an isolated laminin 10. Laminin 10 has vulnerary activity. Laminins  
 CC are useful in maintaining cell/tissue phenotype as well as promoting  
 CC cell growth and differentiation in tissue repair development.  
 CC Specifically, laminin 10 can be used for accelerating the healing  
 CC injuries of vascular tissue, improving the biocompatibility of grafts  
 CC useful for treating such injuries, for promoting re-endothelialisation  
 CC at the site of vascular injuries, and promote cell attachment and  
 CC subsequent cell stasis, proliferation, differentiation, and/or  
 CC migration. The present sequence represents the 2743 N-terminal amino acid  
 CC sequence of human laminin alpha 5, which is used in the exemplification  
 CC of the present invention.  
 XX  
 SQ Sequence 2743 AA;

Query Match 81.2%; Score 39; DB 23; Length 2743;  
 Best Local Similarity 100.0%; Pred. No. 6.9e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVLRGGR 7  
 |||||  
 Db 1928 CVLRGGR 1934

# RESULT 12

ABB81588

ID ABB81588 standard; Protein; 3695 AA.

XX

AC ABB81588;

XX

DT 19-SEP-2002 (first entry)

XX

DE Human laminin alpha 5 protein SEQ ID NO:2.

XX

KW Laminin alpha 5; laminin 10; vulnerary; cell growth; differentiation;  
 KW tissue repair development; laminin; healing; vascular tissue;  
 KW re-endothelialisation; vascular injury; cell attachment; cell stasis;  
 KW proliferation; migration.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Peptide 1..35

FT /label= signal

FT Protein 36..3695

FT /label= laminin\_alpha\_5

XX  
 PN WO200250111-A2.  
 XX  
 PD 27-JUN-2002.  
 XX  
 PF 21-DEC-2001; 2001WO-US51035.  
 XX  
 PR 21-DEC-2000; 2000US-257449P.  
 PR 28-MAR-2001; 2001US-279282P.  
 PR 13-NOV-2001; 2001US-0279282.  
 XX  
 PA (BIOS-) BIOSTRATUM INC.  
 XX  
 PI Tryggvason K, Doi M, Thyboll J;  
 XX  
 DR WPI; 2002-557650/59.  
 DR N-PSDB; ABQ72906.  
 XX  
 PT New human laminin-10 proteins, useful for accelerating the healing of  
 PT vascular tissue, improving the biocompatibility of grafts, or for  
 PT promoting re-endothelialization at the site of vascular injuries -  
 XX  
 PS Claim 5; Page 68-79; 231pp; English.  
 XX  
 CC The present sequence represents human laminin alpha 5. Also described  
 CC is an isolated laminin 10. Laminin 10 has vulnerary activity. Laminins  
 CC are useful in maintaining cell/tissue phenotype as well as promoting  
 CC cell growth and differentiation in tissue repair development.  
 CC Specifically, laminin 10 can be used for accelerating the healing  
 CC injuries of vascular tissue, improving the biocompatibility of grafts  
 CC useful for treating such injuries, for promoting re-endothelialisation  
 CC at the site of vascular injuries, and promote cell attachment and  
 CC subsequent cell stasis, proliferation, differentiation, and/or  
 CC migration.  
 XX  
 SQ Sequence 3695 AA;

Query Match 81.2%; Score 39; DB 23; Length 3695;  
 Best Local Similarity 100.0%; Pred. No. 8.9e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVLRGGR 7  
 |||||  
 Db 1928 CVLRGGR 1934

# RESULT 13

AAE17310

ID AAE17310 standard; Protein; 3696 AA.

XX

AC AAE17310;

XX

DT 18-APR-2002 (first entry)

XX

DE Human laminin alpha protein, sbg417005LAMININ\_ALPHA #2.

XX

KW Human; therapy; wound healing disorder; vaccine; cancer; infection;

KW autoimmune disorder; haematopoietic disorder; inflammation; arthritis;  
KW Parkinson's disease; Huntington's chorea; schizophrenia; antiarrhythmic;  
KW multiple sclerosis; Alzheimer's disease; analgesic; cardiant; asthma;  
KW ischaemia; stroke; AIDS; bone disease; atherosclerosis; brain disorder;  
KW depression; cardiovascular disease; myocardial infarction; renal failure;  
KW respiratory disease; liver disorder; Fanconi's syndrome; spleen disorder;  
KW type II diabetes mellitus; skeletal muscle disorder; immunosuppressive;  
KW hypersplenism; renal disease; hypoglycaemia; gastrointestinal disease;  
KW nootropic; cirrhosis; Hodgkin's disease; neuroleptic; antiinflammatory;  
KW haemostatic; vulnerary; anticonvulsant; antirheumatic; neuroprotective;  
KW nephrotropic; hypotensive; vasotropic; cytostatic; cerebroprotective;  
KW allergy; laminin alpha protein.

XX

OS Homo sapiens.

XX

PN WO200198342-A1.

XX

PD 27-DEC-2001.

XX

PF 22-JUN-2001; 2001WO-US19929.

XX

PR 22-JUN-2000; 2000US-213156P.

PR 22-JUN-2000; 2000US-213161P.

XX

PA (SMIK ) SMITHKLINE BEECHAM CORP.

PA (SMIK ) SMITHKLINE BEECHAM PLC.

PA (GLAX ) GLAXO GROUP LTD.

XX

PI Agarwal P, Cogswell JP, Kabnic KS, Lai Y, Martensen SA;

PI Murdock PR, Smith RF, Strum JC, Xiang Z, Xie Q, Rizni SK;

XX

DR WPI; 2002-139783/18.

DR N-PSDB; AAD27805.

XX

PT Novel secreted and membrane-associated polypeptides and polynucleotides  
PT useful for preventing, ameliorating or correcting dysfunction or  
PT disease including diabetes, cancer, hypertension and growth  
PT abnormalities -

XX

PS Claim 1; Page 115-122; 138pp; English.

XX

CC The invention relates to secreted and membrane-associated polypeptides  
CC and polynucleotides. The sequences of the invention are useful in  
CC diagnostic assays for detecting diseases associated with inappropriate  
CC activity or levels of these polynucleotides, and in identifying their  
CC agonists and antagonists that are potentially useful in therapy. The  
CC sequences of the invention are useful as vaccines for inducing  
CC immunological response. The sequences of the invention are useful for  
CC treating cancers, infections, autoimmune disorders, haematopoietic  
CC disorders, wound healing disorders, cholesteryl ester storage disease,  
CC inflammation, congenital muscular dystrophy, junctional epidermolysis  
CC bullosa, Parkinson's disease, Huntington's chorea, multiple sclerosis,  
CC viral and bacterial infections, Alzheimer's disease, asthma, arthritis,  
CC allergies, schizophrenia, sbg442445PR0a-associated disorders,  
CC septicaemia, psoriasis, inflammatory bowel disease, transplant rejection,  
CC graft versus host disease, ischaemia, stroke, acute respiratory disease  
CC syndrome, restenosis, brain injury, AIDS, bone diseases, atherosclerosis,

CC brain disorders including parasupranuclear palsy, myotonic dystrophy,  
 CC depression, anxiety disorders and sleep disorders, cardiovascular  
 CC diseases including congestive heart failure and myocardial infarction,  
 CC respiratory diseases including chronic obstructive pulmonary disease,  
 CC acute bronchitis and adult respiratory distress syndrome, liver disorders  
 CC including hypercholesterolaemia, hypertriglyceridaemia, cirrhosis, viral  
 CC and non-viral hepatitis, type II diabetes mellitus, renal disease  
 CC including acute and chronic renal failure, glomerulonephritis, Fanconi's  
 CC syndrome, cystinuria, skeletal muscle disorders including hypoglycaemia  
 CC and tendinitis, gastrointestinal diseases including intestinal  
 CC obstruction and tropical sprue, spleen disorders including hypersplenism,  
 CC Hodgkin's disease and malignant lymphoma, testicular cancer, male  
 CC reproductive diseases including low testosterone and male infertility.  
 CC The present sequence is human laminin alpha protein.  
 XX  
 SQ Sequence 3696 AA;

Query Match 81.2%; Score 39; DB 23; Length 3696;  
 Best Local Similarity 100.0%; Pred. No. 8.9e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVLRGGR 7  
 |||||  
 Db 1928 CVLRGGR 1934

#### RESULT 14

AAE17309

ID AAE17309 standard; Protein; 3705 AA.

XX

AC AAE17309;

XX

DT 18-APR-2002 (first entry)

XX

DE Human laminin alpha protein, sbg417005LAMININ\_ALPHA #1.

XX

KW Human; therapy; wound healing disorder; vaccine; cancer; infection;  
 KW autoimmune disorder; haematopoietic disorder; inflammation; arthritis;  
 KW Parkinson's disease; Huntington's chorea; schizophrenia; antiarrhythmic;  
 KW multiple sclerosis; Alzheimer's disease; analgesic; cardiant; asthma;  
 KW ischaemia; stroke; AIDS; bone disease; atherosclerosis; brain disorder;  
 KW depression; cardiovascular disease; myocardial infarction; renal failure;  
 KW respiratory disease; liver disorder; Fanconi's syndrome; spleen disorder;  
 KW type II diabetes mellitus; skeletal muscle disorder; immunosuppressive;  
 KW hypersplenism; renal disease; hypoglycaemia; gastrointestinal disease;  
 KW nootropic; cirrhosis; Hodgkin's disease; neuroleptic; antiinflammatory;  
 KW haemostatic; vulnerary; anticonvulsant; antirheumatic; neuroprotective;  
 KW nephrotropic; hypotensive; vasotropic; cytostatic; cerebroprotective;  
 KW allergy; laminin alpha protein.

XX

OS Homo sapiens.

XX

PN WO200198342-A1.

XX

PD 27-DEC-2001.

XX

PF 22-JUN-2001; 2001WO-US19929.

XX  
PR 22-JUN-2000; 2000US-213156P.  
PR 22-JUN-2000; 2000US-213161P.  
XX  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
PA (SMIK ) SMITHKLINE BEECHAM PLC.  
PA (GLAX ) GLAXO GROUP LTD.  
XX  
PI Agarwal P, Cogswell JP, Kabnic KS, Lai Y, Martensen SA;  
PI Murdock PR, Smith RF, Strum JC, Xiang Z, Xie Q, Rizni SK;  
XX  
DR WPI; 2002-139783/18.  
DR N-PSDB; AAD27804.  
XX  
PT Novel secreted and membrane-associated polypeptides and polynucleotides  
PT useful for preventing, ameliorating or correcting dysfunction or  
PT disease including diabetes, cancer, hypertension and growth  
PT abnormalities -  
XX  
PS Claim 1; Page 107-114; 138pp; English.  
XX  
CC The invention relates to secreted and membrane-associated polypeptides  
CC and polynucleotides. The sequences of the invention are useful in  
CC diagnostic assays for detecting diseases associated with inappropriate  
CC activity or levels of these polynucleotides, and in identifying their  
CC agonists and antagonists that are potentially useful in therapy. The  
CC sequences of the invention are useful as vaccines for inducing  
CC immunological response. The sequences of the invention are useful for  
CC treating cancers, infections, autoimmune disorders, haematopoietic  
CC disorders, wound healing disorders, cholesteryl ester storage disease,  
CC inflammation, congenital muscular dystrophy, junctional epidermolysis  
CC bullosa, Parkinson's disease, Huntington's chorea, multiple sclerosis,  
CC viral and bacterial infections, Alzheimer's disease, asthma, arthritis,  
CC allergies, schizophrenia, sbg442445PROa-associated disorders,  
CC septicaemia, psoriasis, inflammatory bowel disease, transplant rejection,  
CC graft versus host disease, ischaemia, stroke, acute respiratory disease  
CC syndrome, restenosis, brain injury, AIDS, bone diseases, atherosclerosis,  
CC brain disorders including parasupranuclear palsy, myotonic dystrophy,  
CC depression, anxiety disorders and sleep disorders, cardiovascular  
CC diseases including congestive heart failure and myocardial infarction,  
CC respiratory diseases including chronic obstructive pulmonary disease,  
CC acute bronchitis and adult respiratory distress syndrome, liver disorders  
CC including hypercholesterolaemia, hypertriglyceridaemia, cirrhosis, viral  
CC and non-viral hepatitis, type II diabetes mellitus, renal disease  
CC including acute and chronic renal failure, glomerulonephritis, Fanconi's  
CC syndrome, cystinuria, skeletal muscle disorders including hypoglycaemia  
CC and tendinitis, gastrointestinal diseases including intestinal  
CC obstruction and tropical sprue, spleen disorders including hypersplenism,  
CC Hodgkin's disease and malignant lymphoma, testicular cancer, male  
CC reproductive diseases including low testosterone and male infertility.  
CC The present sequence is human laminin alpha protein.  
XX  
SQ Sequence 3705 AA;

Query Match 81.2%; Score 39; DB 23; Length 3705;  
Best Local Similarity 100.0%; Pred. No. 8.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy            1 CVLRGGR 7  
             |||||  
Db           1928 CVLRGGR 1934

RESULT 15

AA017769

ID    AA017769 standard; peptide; 8 AA.

XX

AC    AA017769;

XX

DT    30-AUG-2002    (first entry)

XX

DE    Human beta-defensin-3 derivative #4.

XX

KW    Human; beta-defensin-3; hBD-3; bacterial infection; gene therapy;  
KW    respiratory system; cystic fibrosis; inflammation; urogenital tract;  
KW    antibacterial; fungicide; cytostatic; antiinflammatory; antiulcer;  
KW    gastrointestinal tract; septicemia; apoptosis induction; cancer.

XX

OS    Homo sapiens.

XX

FH    Key                            Location/Qualifiers

FT    Modified-site            1

FT                                        /note= "may be bound to between 1 and 50 amino acids"

FT    Modified-site            8

FT                                        /note= "may be bound to between 1 and 50 amino acids"

XX

PN    WO200240512-A2.

XX

PD    23-MAY-2002.

XX

PF    14-NOV-2001; 2001WO-EP13174.

XX

PR    14-NOV-2000; 2000DE-1056365.

PR    30-MAR-2001; 2001DE-1016220.

XX

PA    (IPFP-) IPF PHARM GMBH.

XX

PI    Forssmann W, Kluever E, Conejo-Garcia J, Adermann K, Bals R;  
PI    Maegert H;

XX

DR    WPI; 2002-435959/46.

XX

PT    New human beta-defensin 3, useful for treating or preventing microbial  
PT    infection and tumors, also related nucleic acid -

XX

PS    Claim 3; Page 24; 36pp; German.

XX

CC    The present invention relates to human beta-defensin-3 (hBD-3) and its  
CC    derivatives. The peptide, its coding sequence and vectors containing the  
CC    coding sequence are useful in (gene) therapy and diagnosis, especially  
CC    for preventing or treating a wide range of microbial infections  
CC    (particularly Burkholderia cepacia and Pseudomonas aeruginosa in the  
CC    respiratory tract, especially in cases of cystic fibrosis, and  
CC    Helicobacter pylori, also inflammatory diseases of the gastrointestinal

CC and urogenital tracts, sepsis and yeast infections), and for inducing  
CC apoptosis for treating malignant melanoma and tumours. The present  
CC sequence is a derivative of human BD-3.

XX

SQ Sequence 8 AA;

Query Match 79.2%; Score 38; DB 23; Length 8;  
Best Local Similarity 75.0%; Pred. No. 9.3e+05;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
| :|||||  
Db 1 CRVRGGRC 8

Search completed: November 13, 2003, 09:45:24  
Job time : 27.9167 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:45:35 ; Search time 16.5833 Seconds  
(without alignments)  
88.069 Million cell updates/sec

Title: US-09-228-866-4  
Perfect score: 48  
Sequence: 1 CVLRGGRC 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*



11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*  
 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*  
 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*  
 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*  
 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	%		DB	ID	Description
		Query Match	Length			
1	41	85.4	36	12	US-10-289-660-184	Sequence 184, Appl
2	41	85.4	36	15	US-10-133-128-184	Sequence 184, Appl
3	41	85.4	4123	15	US-10-213-509-5	Sequence 5, Appli
4	39	81.2	35	12	US-10-252-734-8	Sequence 8, Appli
5	39	81.2	2743	15	US-10-037-182-36	Sequence 36, Appl
6	39	81.2	3695	15	US-10-037-182-2	Sequence 2, Appli
7	38	79.2	34	12	US-10-091-166B-64	Sequence 64, Appl
8	38	79.2	34	12	US-10-272-121-64	Sequence 64, Appl
9	38	79.2	34	12	US-10-409-366-64	Sequence 64, Appl
10	38	79.2	34	12	US-10-409-532-64	Sequence 64, Appl
11	38	79.2	35	12	US-10-091-166B-62	Sequence 62, Appl
12	38	79.2	35	12	US-10-091-166B-63	Sequence 63, Appl
13	38	79.2	35	12	US-10-272-121-62	Sequence 62, Appl
14	38	79.2	35	12	US-10-272-121-63	Sequence 63, Appl
15	38	79.2	35	12	US-10-409-366-62	Sequence 62, Appl
16	38	79.2	35	12	US-10-409-366-63	Sequence 63, Appl
17	38	79.2	35	12	US-10-409-532-62	Sequence 62, Appl
18	38	79.2	35	12	US-10-409-532-63	Sequence 63, Appl
19	38	79.2	35	12	US-10-252-734-7	Sequence 7, Appli
20	38	79.2	36	12	US-10-091-166B-60	Sequence 60, Appl
21	38	79.2	36	12	US-10-091-166B-61	Sequence 61, Appl
22	38	79.2	36	12	US-10-272-121-60	Sequence 60, Appl
23	38	79.2	36	12	US-10-272-121-61	Sequence 61, Appl
24	38	79.2	36	12	US-10-409-366-60	Sequence 60, Appl
25	38	79.2	36	12	US-10-409-366-61	Sequence 61, Appl
26	38	79.2	36	12	US-10-409-532-60	Sequence 60, Appl
27	38	79.2	36	12	US-10-409-532-61	Sequence 61, Appl
28	38	79.2	37	12	US-10-091-166B-58	Sequence 58, Appl
29	38	79.2	37	12	US-10-091-166B-59	Sequence 59, Appl
30	38	79.2	37	12	US-10-272-121-58	Sequence 58, Appl
31	38	79.2	37	12	US-10-272-121-59	Sequence 59, Appl
32	38	79.2	37	12	US-10-409-366-58	Sequence 58, Appl
33	38	79.2	37	12	US-10-409-366-59	Sequence 59, Appl
34	38	79.2	37	12	US-10-409-532-58	Sequence 58, Appl
35	38	79.2	37	12	US-10-409-532-59	Sequence 59, Appl
36	38	79.2	38	12	US-10-091-166B-56	Sequence 56, Appl
37	38	79.2	38	12	US-10-091-166B-57	Sequence 57, Appl
38	38	79.2	38	12	US-10-272-121-56	Sequence 56, Appl
39	38	79.2	38	12	US-10-272-121-57	Sequence 57, Appl

40	38	79.2	38	12	US-10-409-366-56	Sequence 56, Appl
41	38	79.2	38	12	US-10-409-366-57	Sequence 57, Appl
42	38	79.2	38	12	US-10-409-532-56	Sequence 56, Appl
43	38	79.2	38	12	US-10-409-532-57	Sequence 57, Appl
44	38	79.2	39	12	US-10-091-166B-54	Sequence 54, Appl
45	38	79.2	39	12	US-10-091-166B-55	Sequence 55, Appl

# ALIGNMENTS

## RESULT 1

US-10-289-660-184

```
; Sequence 184, Application US/10289660
; Publication No. US20030157561A1
; GENERAL INFORMATION:
; APPLICANT: KOLKMAN, JOOST A.
; APPLICANT: STEMMER, WILLEM P.C.
; APPLICANT: GOVINDARAJAN, SRIDHAR
; TITLE OF INVENTION: COMBINATORIAL LIBRARIES OF MONOMER DOMAINS
; FILE REFERENCE: 0319.510US
; CURRENT APPLICATION NUMBER: US/10/289,660
; CURRENT FILING DATE: 2003-11-06
; PRIOR APPLICATION NUMBER: 10/133,128
; PRIOR FILING DATE: 2002-04-26
; PRIOR APPLICATION NUMBER: 60/374,107
; PRIOR FILING DATE: 2002-04-18
; PRIOR APPLICATION NUMBER: 60/333,359
; PRIOR FILING DATE: 2001-11-26
; PRIOR APPLICATION NUMBER: 60/337,209
; PRIOR FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: 60/286,823
; PRIOR FILING DATE: 2001-04-26
; NUMBER OF SEQ ID NOS: 244
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 184
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-289-660-184
```

```
Query Match          85.4%;  Score 41;  DB 12;  Length 36;
Best Local Similarity 87.5%;  Pred. No. 4;
Matches      7;  Conservative    0;  Mismatches    1;  Indels      0;  Gaps      0;
```

```
Qy      1 CVLRGGRC 8
        ||||| |
Db      14 CVLRGGPC 21
```

## RESULT 2

US-10-133-128-184

```
; Sequence 184, Application US/10133128
; Publication No. US20030082630A1
; GENERAL INFORMATION:
; APPLICANT: KOLKMAN, JOOST A.
; APPLICANT: STEMMER, WILLEM P.C.
```

```

; TITLE OF INVENTION: COMBINATORIAL LIBRARIES OF MONOMER DOMAINS
; FILE REFERENCE: 0319.410US
; CURRENT APPLICATION NUMBER: US/10/133,128
; CURRENT FILING DATE: 2002-04-26
; PRIOR APPLICATION NUMBER: 60/374,107
; PRIOR FILING DATE: 2002-04-18
; PRIOR APPLICATION NUMBER: 60/333,359
; PRIOR FILING DATE: 2001-11-26
; PRIOR APPLICATION NUMBER: 60/337,209
; PRIOR FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: 60/286,823
; PRIOR FILING DATE: 2001-04-26
; NUMBER OF SEQ ID NOS: 244
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 184
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-133-128-184

```

```

Query Match          85.4%;  Score 41;  DB 15;  Length 36;
Best Local Similarity 87.5%;  Pred. No. 4;
Matches      7;  Conservative    0;  Mismatches    1;  Indels      0;  Gaps      0;

```

```

Qy          1 CVLRGGRC 8
             ||||| |
Db          14 CVLRGGPC 21

```

# RESULT 3

```

US-10-213-509-5
; Sequence 5, Application US/10213509
; Publication No. US20030054485A1
; GENERAL INFORMATION:
; APPLICANT: Weiss, Joseph
; APPLICANT: Scott, Matthew
; TITLE OF INVENTION: JELLY BELLY GENES AND THEIR USES
; FILE REFERENCE: STAN-232
; CURRENT APPLICATION NUMBER: US/10/213,509
; CURRENT FILING DATE: 2002-08-06
; PRIOR APPLICATION NUMBER: 60/311,720
; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 4123
; TYPE: PRT
; ORGANISM: H. sapiens
US-10-213-509-5

```

```

Query Match          85.4%;  Score 41;  DB 15;  Length 4123;
Best Local Similarity 87.5%;  Pred. No. 2.9e+02;
Matches      7;  Conservative    0;  Mismatches    1;  Indels      0;  Gaps      0;

```

```

Qy          1 CVLRGGRC 8
             ||||| |
Db          2221 CVLRGGPC 2228

```

RESULT 4

US-10-252-734-8

; Sequence 8, Application US/10252734  
; Publication No. US20030176652A1  
; GENERAL INFORMATION:  
; APPLICANT: MCCRAY, JR., PAUL B.  
; APPLICANT: SCHUTTE, BRIAN C.  
; APPLICANT: JIA, HONG PENG  
; APPLICANT: CASAVANT, THOMAS L.  
; TITLE OF INVENTION: HUMAN AND MOUSE b-DEFENSINS, ANTIMICROBIAL PEPTIDES  
; FILE REFERENCE: IOWA:041US  
; CURRENT APPLICATION NUMBER: US/10/252,734  
; CURRENT FILING DATE: 2002-09-23  
; PRIOR APPLICATION NUMBER: 60/323,991  
; PRIOR FILING DATE: 2001-09-21  
; NUMBER OF SEQ ID NOS: 82  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 35  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-10-252-734-8

Query Match 81.2%; Score 39; DB 12; Length 35;  
Best Local Similarity 75.0%; Pred. No. 8.4;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
| :|||||  
Db 3 CRIRGGRC 10

RESULT 5

US-10-037-182-36

; Sequence 36, Application US/10037182  
; Publication No. US20030044899A1  
; GENERAL INFORMATION:  
; APPLICANT: Tryggvason, Karl  
; APPLICANT: Doi, Masayuki  
; APPLICANT: Thyboll, Jill  
; TITLE OF INVENTION: Recombinant Laminin 10  
; FILE REFERENCE: 99-274-F  
; CURRENT APPLICATION NUMBER: US/10/037,182  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 60/257,449  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/279,282  
; PRIOR FILING DATE: 2001-03-28  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 36  
; LENGTH: 2743  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-037-182-36

Query Match 81.2%; Score 39; DB 15; Length 2743;  
Best Local Similarity 100.0%; Pred. No. 4.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVLRGGR 7  
|||  
Db 1928 CVLRGGR 1934

RESULT 6

US-10-037-182-2  
; Sequence 2, Application US/10037182  
; Publication No. US20030044899A1  
; GENERAL INFORMATION:  
; APPLICANT: Tryggvason, Karl  
; APPLICANT: Doi, Masayuki  
; APPLICANT: Thyboll, Jill  
; TITLE OF INVENTION: Recombinant Laminin 10  
; FILE REFERENCE: 99-274-F  
; CURRENT APPLICATION NUMBER: US/10/037,182  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 60/257,449  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/279,282  
; PRIOR FILING DATE: 2001-03-28  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 3695  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-037-182-2

Query Match 81.2%; Score 39; DB 15; Length 3695;  
Best Local Similarity 100.0%; Pred. No. 5.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVLRGGR 7  
|||  
Db 1928 CVLRGGR 1934

RESULT 7

US-10-091-166B-64  
; Sequence 64, Application US/10091166B  
; Publication No. US20030143671A1  
; GENERAL INFORMATION:  
; APPLICANT: Adler, David A.  
; APPLICANT: Holloway, James L.  
; APPLICANT: Baidur, Nand  
; APPLICANT: Beigel-Orme, Stephanie  
; APPLICANT: Sheppard, Paul O.  
; TITLE OF INVENTION: NOVEL BETA-DEFENSINS  
; FILE REFERENCE: 97-44D1  
; CURRENT APPLICATION NUMBER: US/10/091,166B  
; CURRENT FILING DATE: 2002-03-05

```

; PRIOR APPLICATION NUMBER: US 09/636,399
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: US 09/344,097
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/150,786
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: US 60/064,294
; PRIOR FILING DATE: 1997-11-05
; PRIOR APPLICATION NUMBER: US 60/058,335
; PRIOR FILING DATE: 1997-09-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 64
; LENGTH: 34
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Defensin polypeptide
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (31)...(31)
; OTHER INFORMATION: leucine, isoleucine, valine, phenylalanine, or
; OTHER INFORMATION: methionine
US-10-091-166B-64

```

```

Query Match          79.2%; Score 38; DB 12; Length 34;
Best Local Similarity 75.0%; Pred. No. 12;
Matches      6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 CVLRGGRC 8
        | :|||||
Db      1 CRVRGGRC 8

```

```

RESULT 8
US-10-272-121-64
; Sequence 64, Application US/10272121
; Publication No. US20030157638A1
; GENERAL INFORMATION:
; APPLICANT: Adler, David A.
; APPLICANT: Holloway, James L.
; APPLICANT: Baindur, Nand
; APPLICANT: Beigel-Orme, Stephanie
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: NOVEL BETA-DEFENSINS
; FILE REFERENCE: 97-44D2
; CURRENT APPLICATION NUMBER: US/10/272,121
; CURRENT FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: US 09/636,399
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: US 09/344,097
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/150,786
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: US 60/064,294
; PRIOR FILING DATE: 1997-11-05
; PRIOR APPLICATION NUMBER: US 60/058,335

```

```
; PRIOR FILING DATE: 1997-09-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 64
; LENGTH: 34
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Defensin polypeptide
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (31)...(31)
; OTHER INFORMATION: leucine, isoleucine, valine, phenylalanine, or
; OTHER INFORMATION: methionine
US-10-272-121-64
```

```
Query Match          79.2%;  Score 38;  DB 12;  Length 34;
Best Local Similarity 75.0%;  Pred. No. 12;
Matches      6;  Conservative    1;  Mismatches    1;  Indels      0;  Gaps      0;
```

```
Qy      1 CVLRGGRC 8
        | :|||||
Db      1 CRVRGGRC 8
```

#### RESULT 9

```
US-10-409-366-64
; Sequence 64, Application US/10409366
; Publication No. US20030166912A1
; GENERAL INFORMATION:
; APPLICANT: Adler, David A.
; APPLICANT: Holloway, James L.
; APPLICANT: Baindur, Nand
; APPLICANT: Beigel-Orme, Stephanie
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: NOVEL BETA-DEFENSINS
; FILE REFERENCE: 97-44C2
; CURRENT APPLICATION NUMBER: US/10/409,366
; CURRENT FILING DATE: 2003-04-07
; PRIOR APPLICATION NUMBER: US/09/636,399A
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/058,335
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/064,294
; PRIOR FILING DATE: 1997-11-05
; PRIOR APPLICATION NUMBER: 09/150,786
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 09/636,399
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 64
; LENGTH: 34
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Defensin polypeptide
```

```
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (31)...(31)
; OTHER INFORMATION: Xaa is Ile, Leu, Val, Phe, or Met
US-10-409-366-64
```

```
Query Match          79.2%;  Score 38;  DB 12;  Length 34;
Best Local Similarity 75.0%;  Pred. No. 12;
Matches      6;  Conservative    1;  Mismatches    1;  Indels      0;  Gaps      0;
```

```
Qy      1 CVLRGGRC 8
        | :|||||
Db      1 CRVRGGRC 8
```

RESULT 10

```
US-10-409-532-64
; Sequence 64, Application US/10409532
; Publication No. US20030166913A1
; GENERAL INFORMATION:
; APPLICANT: Adler, David A.
; APPLICANT: Holloway, James L.
; APPLICANT: Baindur, Nand
; APPLICANT: Beigel-Orme, Stephanie
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: NOVEL BETA-DEFENSINS
; FILE REFERENCE: 97-44C2
; CURRENT APPLICATION NUMBER: US/10/409,532
; CURRENT FILING DATE: 2003-04-07
; PRIOR APPLICATION NUMBER: US/09/636,399A
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/058,335
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/064,294
; PRIOR FILING DATE: 1997-11-05
; PRIOR APPLICATION NUMBER: 09/150,786
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 09/636,399
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 64
; LENGTH: 34
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Defensin polypeptide
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (31)...(31)
; OTHER INFORMATION: Xaa is Ile, Leu, Val, Phe, or Met
US-10-409-532-64
```

```
Query Match          79.2%;  Score 38;  DB 12;  Length 34;
Best Local Similarity 75.0%;  Pred. No. 12;
Matches      6;  Conservative    1;  Mismatches    1;  Indels      0;  Gaps      0;
```



Qy 1 CVLRGGRC 8  
| :|||||  
Db 1 CRVRGGRC 8

RESULT 11

US-10-091-166B-62

; Sequence 62, Application US/10091166B  
; Publication No. US20030143671A1  
; GENERAL INFORMATION:  
; APPLICANT: Adler, David A.  
; APPLICANT: Holloway, James L.  
; APPLICANT: Baindur, Nand  
; APPLICANT: Beigel-Orme, Stephanie  
; APPLICANT: Sheppard, Paul O.  
; TITLE OF INVENTION: NOVEL BETA-DEFENSINS  
; FILE REFERENCE: 97-44D1  
; CURRENT APPLICATION NUMBER: US/10/091,166B  
; CURRENT FILING DATE: 2002-03-05  
; PRIOR APPLICATION NUMBER: US 09/636,399  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: US 09/344,097  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: US 09/150,786  
; PRIOR FILING DATE: 1998-09-10  
; PRIOR APPLICATION NUMBER: US 60/064,294  
; PRIOR FILING DATE: 1997-11-05  
; PRIOR APPLICATION NUMBER: US 60/058,335  
; PRIOR FILING DATE: 1997-09-10  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 62  
; LENGTH: 35  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Defensin polypeptide  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (32)...(32)  
; OTHER INFORMATION: leucine, isoleucine, valine, phenylalanine, or  
; OTHER INFORMATION: methionine  
US-10-091-166B-62

Query Match 79.2%; Score 38; DB 12; Length 35;  
Best Local Similarity 75.0%; Pred. No. 12;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
| :|||||  
Db 2 CRVRGGRC 9

RESULT 12

US-10-091-166B-63

; Sequence 63, Application US/10091166B  
; Publication No. US20030143671A1

```

; GENERAL INFORMATION:
; APPLICANT: Adler, David A.
; APPLICANT: Holloway, James L.
; APPLICANT: Baindur, Nand
; APPLICANT: Beigel-Orme, Stephanie
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: NOVEL BETA-DEFENSINS
; FILE REFERENCE: 97-44D1
; CURRENT APPLICATION NUMBER: US/10/091,166B
; CURRENT FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 09/636,399
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: US 09/344,097
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/150,786
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: US 60/064,294
; PRIOR FILING DATE: 1997-11-05
; PRIOR APPLICATION NUMBER: US 60/058,335
; PRIOR FILING DATE: 1997-09-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 63
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Defensin polypeptide
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (31)...(31)
; OTHER INFORMATION: leucine, isoleucine, valine, phenylalanine, or
; OTHER INFORMATION: methionine
US-10-091-166B-63

```

```

Query Match          79.2%; Score 38; DB 12; Length 35;
Best Local Similarity 75.0%; Pred. No. 12;
Matches      6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 CVLRGGRC 8
        | :|||||
Db      1 CRVRGGRC 8

```

```

RESULT 13
US-10-272-121-62
; Sequence 62, Application US/10272121
; Publication No. US20030157638A1
; GENERAL INFORMATION:
; APPLICANT: Adler, David A.
; APPLICANT: Holloway, James L.
; APPLICANT: Baindur, Nand
; APPLICANT: Beigel-Orme, Stephanie
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: NOVEL BETA-DEFENSINS
; FILE REFERENCE: 97-44D2
; CURRENT APPLICATION NUMBER: US/10/272,121

```

```

; CURRENT FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: US 09/636,399
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: US 09/344,097
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/150,786
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: US 60/064,294
; PRIOR FILING DATE: 1997-11-05
; PRIOR APPLICATION NUMBER: US 60/058,335
; PRIOR FILING DATE: 1997-09-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 62
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Defensin polypeptide
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (32)...(32)
; OTHER INFORMATION: leucine, isoleucine, valine, phenylalanine, or
; OTHER INFORMATION: methionine
US-10-272-121-62

```

```

Query Match          79.2%; Score 38; DB 12; Length 35;
Best Local Similarity 75.0%; Pred. No. 12;
Matches      6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 CVLRGGRC 8
        | :|||||
Db      2 CRVRGGRC 9

```

```

RESULT 14
US-10-272-121-63
; Sequence 63, Application US/10272121
; Publication No. US20030157638A1
; GENERAL INFORMATION:
; APPLICANT: Adler, David A.
; APPLICANT: Holloway, James L.
; APPLICANT: Baindur, Nand
; APPLICANT: Beigel-Orme, Stephanie
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: NOVEL BETA-DEFENSINS
; FILE REFERENCE: 97-44D2
; CURRENT APPLICATION NUMBER: US/10/272,121
; CURRENT FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: US 09/636,399
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: US 09/344,097
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/150,786
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: US 60/064,294
; PRIOR FILING DATE: 1997-11-05

```

```

; PRIOR APPLICATION NUMBER: US 60/058,335
; PRIOR FILING DATE: 1997-09-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 63
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Defensin polypeptide
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (31)...(31)
; OTHER INFORMATION: leucine, isoleucine, valine, phenylalanine, or
; OTHER INFORMATION: methionine
US-10-272-121-63

```

```

Query Match          79.2%; Score 38; DB 12; Length 35;
Best Local Similarity 75.0%; Pred. No. 12;
Matches      6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 CVLRGGRC 8
        | :|||||
Db      1 CRVRGGRC 8

```

```

RESULT 15
US-10-409-366-62
; Sequence 62, Application US/10409366
; Publication No. US20030166912A1
; GENERAL INFORMATION:
; APPLICANT: Adler, David A.
; APPLICANT: Holloway, James L.
; APPLICANT: Baindur, Nand
; APPLICANT: Beigel-Orme, Stephanie
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: NOVEL BETA-DEFENSINS
; FILE REFERENCE: 97-44C2
; CURRENT APPLICATION NUMBER: US/10/409,366
; CURRENT FILING DATE: 2003-04-07
; PRIOR APPLICATION NUMBER: US/09/636,399A
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/058,335
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/064,294
; PRIOR FILING DATE: 1997-11-05
; PRIOR APPLICATION NUMBER: 09/150,786
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 09/636,399
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 62
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:

```

; OTHER INFORMATION: Defensin polypeptide  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (32)...(32)  
; OTHER INFORMATION: Xaa is Phe, Val, Ile, Leu, or Met  
US-10-409-366-62

Query Match 79.2%; Score 38; DB 12; Length 35;  
Best Local Similarity 75.0%; Pred. No. 12;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
| :|||||  
Db 2 CRVRGGRC 9

Search completed: November 13, 2003, 09:58:27  
Job time : 16.5833 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:38:30 ; Search time 8.33333 Seconds  
(without alignments)  
92.322 Million cell updates/sec

Title: US-09-228-866-4  
Perfect score: 48  
Sequence: 1 CVLRGGRC 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_76:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result		%					
No.	Score	Query	Match	Length	DB	ID	Description
1	38	79.2	681	2	F95885		probable iron ABC
2	37	77.1	110	2	S50991		hypothetical prote
3	36	75.0	224	2	S48671		proliferin - human
4	36	75.0	224	2	A05086		proliferin 1 precu
5	36	75.0	224	2	A23159		proliferin 2 precu
6	36	75.0	224	2	S05648		proliferin 3 - mou
7	36	75.0	437	2	S05478		properdin - mouse
8	36	75.0	485	2	A40751		finger protein MZF
9	36	75.0	522	2	B45268		interleukin-9 rece
10	36	75.0	627	2	D75393		serine proteinase,
11	36	75.0	876	2	B96693		probable receptor
12	36	75.0	1154	2	A86318		protein F15H18.11
13	35	72.9	370	2	T22082		hypothetical prote
14	35	72.9	1080	2	PC4394		DNA-directed DNA p
15	35	72.9	2533	2	T28675		alpha-51D immobili
16	35	72.9	2533	2	T28674		alpha-51D-immobili
17	35	72.9	2543	2	T31687		surface antigen - P
18	34	70.8	61	2	S57815		antimicrobial pept
19	34	70.8	63	2	S57816		antimicrobial pept
20	34	70.8	119	1	JAAO92		chorion class A pr
21	34	70.8	150	2	T03977		hypothetical prote
22	34	70.8	409	2	H71030		hypothetical prote
23	34	70.8	435	2	B72418		conserved hypothet
24	34	70.8	547	2	E82422		anaerobic glycerol
25	33	68.8	162	1	QQBE19		BLRF2 protein - hu
26	33	68.8	518	1	WMBE71		protein-serine/thr
27	33	68.8	562	2	AF3424		choline dehydrogen
28	33	68.8	632	2	T38126		probable electron
29	33	68.8	697	2	H71525		probable outer mem
30	33	68.8	697	2	S48964		hypothetical prote
31	33	68.8	700	2	B81682		conserved hypothet
32	33	68.8	780	2	T03156		ribonucleoside-dip
33	33	68.8	802	2	H86056		hypothetical prote
34	33	68.8	802	2	E91210		hypothetical prote
35	33	68.8	1032	2	S65341		probable membrane
36	33	68.8	1056	1	WMAD12		DNA-directed DNA p
37	33	68.8	1056	1	DJAD51		DNA-directed DNA p
38	33	68.8	1061	1	DJAD12		DNA-directed DNA p
39	33	68.8	1115	2	S40241		G protein-coupled
40	33	68.8	1205	2	S64819		probable membrane
41	33	68.8	1221	2	T18550		reverse gyrase cha
42	33	68.8	1277	2	S70306		hypothetical prote
43	33	68.8	1790	1	MMFFB1		laminin beta-1 cha
44	33	68.8	3635	2	T10053		laminin alpha 5 ch
45	32	66.7	36	2	S66282		defensin beta-1 -

## ALIGNMENTS

RESULT 1  
F95885

probable iron ABC transporter permease protein SMb20364 [imported] -  
 Sinorhizobium meliloti (strain 1021) megaplasmid pSymB  
 C;Species: Sinorhizobium meliloti  
 C;Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
 C;Accession: F95885  
 R;Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernandez-Lucas, I.; Becker, A.; Cowie, A.; Gouzy, J.; Golding, B.; Puhler, A.  
 Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001  
 A;Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endosymbiont Sinorhizobium meliloti.  
 A;Reference number: A95842; MUID:21396508; PMID:11481431  
 A;Accession: F95885  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-681 <KUR>  
 A;Cross-references: GB:AL591985; PIDN:CAC48750.1; PID:g15140223; GSPDB:GN00167  
 A;Experimental source: strain 1021, megaplasmid pSymB  
 R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler, F.; Barnett, M.J.; Becker, A.; Boistard, P.; Bothe, G.; Boutry, M.; Bowser, L.; Buhrmester, J.; Cadieu, E.; Capela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.; Gloux, S.; Godrie, T.; Goffeau, A.; Golding, B.; Gouzy, J.; Gurjal, M.; Hernandez-Lucas, I.; Hong, A.; Huizar, L.; Hyman, R.W.; Jones, T.  
 Science 293, 668-672, 2001  
 A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure, V.; Masuy, D.; Palm, C.; Peck, M.C.; Pohl, T.M.; Portetelle, D.; Purnelle, B.; Ramsperger, U.; Surzycki, R.; Thebault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.C.; Batut, J.  
 A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
 A;Reference number: A96039; MUID:21368234; PMID:11474104  
 A;Contents: annotation  
 C;Genetics:  
 A;Gene: SMb20364  
 A;Genome: plasmid

Query Match 79.2%; Score 38; DB 2; Length 681;  
 Best Local Similarity 75.0%; Pred. No. 28;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
 ||: ||||  
 Db 152 CVVGGGRC 159

RESULT 2  
 S50991  
 hypothetical protein YDR010c - yeast (Saccharomyces cerevisiae)  
 N;Alternate names: hypothetical protein D3209; hypothetical protein PZE110; hypothetical protein YD8119.15c  
 C;Species: Saccharomyces cerevisiae  
 C;Date: 11-Feb-1995 #sequence\_revision 12-May-1995 #text\_change 19-Apr-2002  
 C;Accession: S50991; S63417; S67823; S72108  
 R;Murphy, L.; Richards, C.; Gentles, S.; Harris, D.  
 submitted to the EMBL Data Library, January 1995  
 A;Reference number: S50976  
 A;Accession: S50991

A;Molecule type: DNA  
 A;Residues: 1-110 <MUR>  
 A;Cross-references: EMBL:Z48008; NID:g642799; PIDN:CAA88070.1; PID:g642815;  
 MIPS:YDR010c  
 R;Eide, L.G.; Sander, C.; Prydz, H.  
 submitted to the EMBL Data Library, February 1996  
 A;Description: Sequencing and analysis of a 35.4 kb region on the left arm of  
 chromosome IV for *Saccharomyces cerevisiae* reveal 23 open reading frames.  
 A;Reference number: S63416  
 A;Accession: S63417  
 A;Molecule type: DNA  
 A;Residues: 1-110 <EID>  
 A;Cross-references: EMBL:X95966; NID:g1216215; PIDN:CAA65202.1; PID:g1216217  
 R;Prydz, H.; Eide, L.G.  
 submitted to the Protein Sequence Database, July 1996  
 A;Reference number: S67822  
 A;Accession: S67823  
 A;Molecule type: DNA  
 A;Residues: 1-110 <PRY>  
 A;Cross-references: EMBL:Z74306; NID:g1431427; PIDN:CAA98830.1; PID:g1431428;  
 MIPS:YDR010c  
 A;Experimental source: strain S288C  
 R;Eide, L.G.; Sander, C.; Prydz, H.  
 Yeast 12, 1085-1090, 1996  
 A;Title: Sequencing and analysis of a 35.4 kb region on the left arm of  
 chromosome IV from *Saccharomyces cerevisiae* reveal 23 open reading frames.  
 A;Reference number: S72107; MUID:97051598; PMID:8896275  
 A;Accession: S72108  
 A;Status: nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-110 <EIW>  
 A;Cross-references: EMBL:X95966; NID:g1216215; PIDN:CAA65202.1; PID:g1216217  
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, February  
 1996  
 C;Genetics:  
 A;Cross-references: SGD:S0002417  
 A;Map position: 4R  
 A;Note: YDR010c  
 C;Superfamily: *Saccharomyces* hypothetical protein YDR010c

Query Match 77.1%; Score 37; DB 2; Length 110;  
 Best Local Similarity 85.7%; Pred. No. 10;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VLRGGRC 8  
 |:|||||  
 Db 50 VIRGGRC 56

RESULT 3  
 S48671  
 proliferin - human  
 C;Species: *Homo sapiens* (man)  
 C;Date: 16-Feb-1995 #sequence\_revision 10-Nov-1995 #text\_change 26-Aug-1999  
 C;Accession: S48671  
 R;Gil-Torregrosa, B.; Urdiales, J.L.; Lozano, J.; Mates, J.M.; Sanchez-Jimenez,  
 F.



FEBS Lett. 349, 343-348, 1994  
A;Title: Expression of different mitogen-regulated protein/proliferin mRNAs in Ehrlich carcinoma cells.  
A;Reference number: S48671; MUID:94326948; PMID:8050594  
A;Accession: S48671  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-224 <GIL>  
C;Superfamily: prolactin

Query Match 75.0%; Score 36; DB 2; Length 224;  
Best Local Similarity 62.5%; Pred. No. 27;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
| :| |||  
Db 33 CAMRNGRC 40

#### RESULT 4

A05086  
proliferin 1 precursor - mouse  
N;Alternate names: mitogen regulated protein  
C;Species: Mus musculus (house mouse)  
C;Date: 05-Jun-1987 #sequence\_revision 05-Jun-1987 #text\_change 05-Nov-1999  
C;Accession: A05086; A61095; I48940  
R;Linzer, D.I.H.; Nathans, D.  
Proc. Natl. Acad. Sci. U.S.A. 81, 4255-4259, 1984  
A;Reference number: A05086; MUID:84272617; PMID:6087314  
A;Accession: A05086  
A;Molecule type: mRNA  
A;Residues: 1-224 <LIN>  
A;Cross-references: GB:K02245; NID:g200400; PIDN:AAA39946.1; PID:g200401  
R;Lee, S.J.; Nathans, D.  
Endocrinology 120, 208-213, 1987  
A;Title: Secretion of proliferin.  
A;Reference number: A61095; MUID:87053622; PMID:3780559  
A;Accession: A61095  
A;Molecule type: protein  
A;Residues: 30-32,'X',34-39,'X',41-45 <LEE>  
A;Note: this material was purified from recombinant Chinese hamster ovary cell conditioned medium  
R;Ko, M.S.; Wang, X.; Horton, J.H.; Hagen, M.D.; Takahashi, N.; Maezaki, Y.; Nadeau, J.H.  
Mamm. Genome 5, 349-355, 1994  
A;Title: Genetic mapping of 40 cDNA clones on the mouse genome by PCR.  
A;Reference number: I48934; MUID:94319082; PMID:8043949  
A;Accession: I48940  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 208-224 <RES>  
A;Cross-references: EMBL:U05747; NID:g497086; PIDN:AAB60482.1; PID:g497087  
C;Superfamily: prolactin  
C;Keywords: glycoprotein  
F;1-29/Domain: signal sequence #status predicted <SIG>  
F;30-224/Product: proliferin 1 #status experimental <MAT>

Query Match 75.0%; Score 36; DB 2; Length 224;  
Best Local Similarity 62.5%; Pred. No. 27;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
| :| |||  
Db 33 CAMRNGRC 40

RESULT 5

A23159

proliferin 2 precursor - mouse

C;Species: Mus musculus (house mouse)

C;Date: 29-Aug-1987 #sequence\_revision 29-Aug-1987 #text\_change 26-Aug-1999

C;Accession: A23159

R;Linzer, D.I.H.; Lee, S.J.; Ogren, L.; Talamantes, F.; Nathans, D.

Proc. Natl. Acad. Sci. U.S.A. 82, 4356-4359, 1985

A;Title: Identification of proliferin mRNA and protein in mouse placenta.

A;Reference number: A94049; MUID:85242683; PMID:3859868

A;Accession: A23159

A;Molecule type: mRNA

A;Residues: 1-224 <LIN>

A;Experimental source: BALB/c

C;Superfamily: prolactin

C;Keywords: glycoprotein

F;1-29/Domain: signal sequence #status predicted <SIG>

Query Match 75.0%; Score 36; DB 2; Length 224;  
Best Local Similarity 62.5%; Pred. No. 27;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
| :| |||  
Db 33 CAMRNGRC 40

RESULT 6

S05648

proliferin 3 - mouse

N;Alternate names: mitogen-regulated protein 3

C;Species: Mus musculus (house mouse)

C;Date: 28-Feb-1990 #sequence\_revision 28-Feb-1990 #text\_change 20-Jun-2000

C;Accession: S05648

R;Connor, A.M.; Waterhouse, P.; Khokha, R.; Denhardt, D.T.

Biochim. Biophys. Acta 1009, 75-82, 1989

A;Title: Characterization of a mouse mitogen-regulated protein/proliferin gene and its promoter: a member of the growth hormone/prolactin gene superfamily.

A;Reference number: S05648; MUID:90001249; PMID:2790033

A;Accession: S05648

A;Molecule type: DNA

A;Residues: 1-224 <CON>

A;Cross-references: EMBL:X16009; NID:g53223; PIDN:CAA34146.1; PID:g1200103

C;Genetics:

A;Introns: 11/1; 69/3; 105/3; 166/3

C;Superfamily: prolactin

Query Match 75.0%; Score 36; DB 2; Length 224;

Best Local Similarity 62.5%; Pred. No. 27;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
| : | | |  
Db 33 CAMRNGRC 40

RESULT 7

S05478

properdin - mouse (fragment)

C;Species: Mus musculus (house mouse)

C;Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 17-Nov-2000

C;Accession: S05478

R;Goundis, D.; Reid, K.B.M.

Nature 335, 82-85, 1988

A;Title: Properdin, the terminal complement components, thrombospondin and the circumsporozoite protein of malaria parasites contain similar sequence motifs.

A;Reference number: S05478; MUID:88318954; PMID:3045564

A;Accession: S05478

A;Molecule type: mRNA

A;Residues: 1-437 <GOU>

A;Cross-references: EMBL:X12905; NID:g53786; PIDN:CAA31389.1; PID:g53787

C;Complex: a mixture of homodimers, homotrimers and homotetramers

C;Function:

A;Description: protects C3 convertase (C3bBb) from rapid inactivation

A;Pathway: complement alternate pathway

C;Superfamily: human properdin precursor; thrombospondin type 1 repeat homology

C;Keywords: complement alternate pathway; glycoprotein; homodimer; homotetramer; homotrimer; plasma

F;45-97/Domain: thrombospondin type 1 repeat homology <THR1>

F;104-160/Domain: thrombospondin type 1 repeat homology <THR2>

F;161-224/Domain: thrombospondin type 1 repeat homology <THR3>

F;225-282/Domain: thrombospondin type 1 repeat homology <THR4>

F;283-345/Domain: thrombospondin type 1 repeat homology <THR5>

F;346-408/Domain: thrombospondin type 1 repeat homology <THR6>

F;52,55,108,111,114,165,168,229,232,290,293,350,353,356/Modified site: 2'-mannosyl-tryptophan (Trp) #status predicted

F;366,396/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 75.0%; Score 36; DB 2; Length 437;  
Best Local Similarity 75.0%; Pred. No. 46;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
| | | | : |  
Db 73 CVGRGGQC 80

RESULT 8

A40751

finger protein MZF1 - human

C;Species: Homo sapiens (man)

C;Date: 21-Apr-1992 #sequence\_revision 21-Apr-1992 #text\_change 01-Dec-2000

C;Accession: A40751

R;Hromas, R.; Collins, S.J.; Hickstein, D.; Raskind, W.; Deaven, L.L.; O'Hara, P.; Hagen, F.S.; Kaushansky, K.

J. Biol. Chem. 266, 14183-14187, 1991  
 A;Title: A retinoic acid-responsive human zinc finger gene, MZF-1,  
 preferentially expressed in myeloid cells.  
 A;Reference number: A40751; MUID:91317761; PMID:1860835  
 A;Accession: A40751  
 A;Status: preliminary  
 A;Molecule type: mRNA  
 A;Residues: 1-485 <HRO>  
 A;Cross-references: GB:M58297; NID:g189043; PID:g189044  
 C;Genetics:  
 A;Gene: GDB:ZNF42; MZF-1  
 A;Cross-references: GDB:125898; OMIM:194550  
 A;Map position: 19q13.2-19q13.4  
 C;Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology  
 C;Keywords: DNA binding; transcription regulation; zinc finger

Query Match 75.0%; Score 36; DB 2; Length 485;  
 Best Local Similarity 85.7%; Pred. No. 50;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VLRGGRC 8  
 |:|||||  
 Db 103 VVRGGRC 109

RESULT 9  
 B45268  
 interleukin-9 receptor precursor - human  
 C;Species: Homo sapiens (man)  
 C;Date: 27-Jun-1994 #sequence\_revision 27-Jun-1994 #text\_change 05-Nov-1999  
 C;Accession: B45268  
 R;Renauld, J.C.; Druetz, C.; Kermouni, A.; Houssiau, F.; Uyttenhove, C.; Van  
 Roost, E.; Van Snick, J.  
 Proc. Natl. Acad. Sci. U.S.A. 89, 5690-5694, 1992  
 A;Title: Expression cloning of the murine and human interleukin 9 receptor  
 cDNAs.  
 A;Reference number: A45268; MUID:92302307; PMID:1376929  
 A;Accession: B45268  
 A;Status: preliminary  
 A;Molecule type: mRNA  
 A;Residues: 1-522 <REN>  
 A;Cross-references: GB:M84747; NID:g184508; PIDN:AAA58679.1; PID:g184509  
 C;Keywords: glycoprotein; receptor; T-cell proliferation; transmembrane protein

Query Match 75.0%; Score 36; DB 2; Length 522;  
 Best Local Similarity 62.5%; Pred. No. 53;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
 |:||| |  
 Db 95 CILRGSEC 102

RESULT 10  
 D75393  
 serine proteinase, subtilase family - Deinococcus radiodurans (strain R1)  
 C;Species: Deinococcus radiodurans

C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
 C;Accession: D75393  
 R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.;  
 Dodson, R.J.; Haft, D.H.; Gwinn, M.L.; Nelson, W.C.; Richardson, D.L.; Moffat,  
 K.S.; Qin, H.; Jiang, L.; Pamphile, W.; Crosby, M.; Shen, M.; Vamathevan, J.J.;  
 Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Makarova, K.S.; Aravind, L.;  
 Daly, M.J.; Minton, K.W.; Fleischmann, R.D.; Ketchum, K.A.; Nelson, K.E.;  
 Salzberg, S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 Science 286, 1571-1577, 1999  
 A;Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans*  
 R1.  
 A;Reference number: A75250; MUID:20036896; PMID:10567266  
 A;Accession: D75393  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-627 <WHI>  
 A;Cross-references: GB:AE001990; GB:AE000513; NID:g6459214; PIDN:AAF11026.1;  
 PID:g6459217; TIGR:DR1459; GSPDB:GN00077  
 A;Experimental source: strain R1  
 C;Genetics:  
 A;Gene: DR1459  
 A;Map position: 1

Query Match 75.0%; Score 36; DB 2; Length 627;  
 Best Local Similarity 62.5%; Pred. No. 62;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
 | : ||||  
 Db 487 CAVEGGRC 494

# RESULT 11

B96693  
 probable receptor serine/threonine kinase PR5K T4O24.2 [imported] - *Arabidopsis thaliana*  
 C;Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C;Accession: B96693  
 R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.;  
 Alonso, J.; Altaf, H.; Araujo, R.; Bowman, C.L.; Brooks, S.Y.; Buehler, E.;  
 Chan, A.; Chao, Q.; Chen, H.; Cheuk, R.F.; Chin, C.W.; Chung, M.K.; Conn, L.;  
 Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Dunn, P.; Etgu, P.;  
 Feldblyum, T.V.; Feng, J.; Fong, B.; Fujii, C.Y.; Gill, J.E.; Goldsmith, A.D.;  
 Haas, B.; Hansen, N.F.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000  
 A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.;  
 Kim, C.J.; Koo, H.L.; Kremenetskaia, I.; Kurtz, D.B.; Kwan, A.; Lam, B.; Langin-  
 Hooper, S.; Lee, A.; Lee, J.M.; Lenz, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu,  
 S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, A.; Militscher, J.; Miranda,  
 M.; Nguyen, M.; Nierman, W.C.; Osborne, B.I.; Pai, G.; Peterson, J.; Pham, P.K.;  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.;  
 Tallon, L.J.; Tambunga, G.; Toriumi, M.J.; Town, C.D.; Utterback, T.; van Aken,  
 S.; Vaysberg, M.; Vysotskaia, V.S.; Walker, M.; Wu, D.; Yu, G.; Fraser, C.M.;  
 Venter, J.C.; Davis, R.W.  
 A;Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.

A;Reference number: A86141; MUID:21016719; PMID:11130712  
A;Accession: B96693  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-876 <STO>  
A;Cross-references: GB:AE005173; NID:g11128393; PIDN:AAG31198.1; GSPDB:GN00141  
C;Genetics:  
A;Gene: T4O24.2  
A;Map position: 1

Query Match 75.0%; Score 36; DB 2; Length 876;  
Best Local Similarity 75.0%; Pred. No. 81;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
||| |||  
Db 448 CVLSGGSC 455

#### RESULT 12

A86318

protein F15H18.11 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C;Accession: A86318

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.;  
Alonso, J.; Altaf, H.; Araujo, R.; Bowman, C.L.; Brooks, S.Y.; Buehler, E.;  
Chan, A.; Chao, Q.; Chen, H.; Cheuk, R.F.; Chin, C.W.; Chung, M.K.; Conn, L.;  
Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Dunn, P.; Etgu, P.;  
Feldblyum, T.V.; Feng, J.; Fong, B.; Fujii, C.Y.; Gill, J.E.; Goldsmith, A.D.;  
Haas, B.; Hansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.;  
Kim, C.J.; Koo, H.L.; Kremenetskaia, I.; Kurtz, D.B.; Kwan, A.; Lam, B.; Langin-  
Hooper, S.; Lee, A.; Lee, J.M.; Lenz, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu,  
S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, A.; Militscher, J.; Miranda,  
M.; Nguyen, M.; Nierman, W.C.; Osborne, B.I.; Pai, G.; Peterson, J.; Pham, P.K.;  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.;  
Tallon, L.J.; Tambunga, G.; Toriumi, M.J.; Town, C.D.; Utterback, T.; van Aken,  
S.; Vaysberg, M.; Vysotskaia, V.S.; Walker, M.; Wu, D.; Yu, G.; Fraser, C.M.;  
Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: A86318

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1154 <STO>

A;Cross-references: GB:AE005172; NID:g6714300; PIDN:AAF25996.1; GSPDB:GN00141

C;Genetics:

A;Gene: F15H18.11

A;Map position: 1

Query Match 75.0%; Score 36; DB 2; Length 1154;  
Best Local Similarity 62.5%; Pred. No. 1e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy            1 CVLRGGRC 8  
             |: ||||  
Db           451 CITSGGRC 458

RESULT 13

T22082

hypothetical protein F42A8.1 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C;Accession: T22082

R;Matthews, P.

submitted to the EMBL Data Library, January 1995

A;Reference number: Z19510

A;Accession: T22082

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-370 <WIL>

A;Cross-references: EMBL:Z47809; PIDN:CAA87779.1; GSPDB:GN00020; CESP:F42A8.1

A;Experimental source: clone F42A8

C;Genetics:

A;Gene: CESP:F42A8.1

A;Map position: 2

A;Introns: 33/2; 97/1; 127/3; 169/2; 201/3; 246/3

Query Match                    72.9%;   Score 35;   DB 2;   Length 370;  
Best Local Similarity       62.5%;   Pred. No. 61;  
Matches       5;   Conservative       2;   Mismatches       1;   Indels       0;   Gaps       0;

Qy            1 CVLRGGRC 8  
             |: ||||  
Db           209 CKLQGGKC 216

RESULT 14

PC4394

DNA-directed DNA polymerase (EC 2.7.7.7) - *Ovine adenovirus* OAV287 (fragment)

C;Species: *Ovine adenovirus* OAV287

C;Date: 10-Nov-1997 #sequence\_revision 10-Nov-1997 #text\_change 03-Nov-2000

C;Accession: PC4394

R;Vrati, S.; Brookes, D.E.; Boyle, D.B.; Both, G.W.

Gene 177, 35-41, 1996

A;Title: Nucleotide sequence of ovine adenovirus tripartite leader sequence and homologues of the IVa2, DNA polymerase and terminal proteins.

A;Reference number: JC5648; MUID:97080497; PMID:8921842

A;Accession: PC4394

A;Molecule type: DNA

A;Residues: 1-1080 <VRA>

A;Cross-references: GB:U31557; NID:g1117828; PIDN:AAC55957.1; PID:g1117830

C;Comment: This enzyme is targeted to the nucleus by interaction with the terminal protein precursor which has a nuclear localization.

C;Superfamily: adenovirus DNA-directed DNA polymerase

C;Keywords: nucleotidyltransferase

Query Match                    72.9%;   Score 35;   DB 2;   Length 1080;  
Best Local Similarity       100.0%;   Pred. No. 1.5e+02;  
Matches       6;   Conservative       0;   Mismatches       0;   Indels       0;   Gaps       0;

Qy 3 LRGGRC 8  
 |||||  
Db 545 LRGGRC 550

RESULT 15

T28675

alpha-51D immobilization antigen - Paramecium tetraurelia

C;Species: Paramecium tetraurelia

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 20-Jun-2000

C;Accession: T28675

R;Schwegmann, K.J.

submitted to the EMBL Data Library, March 1996

A;Reference number: Z20506

A;Accession: T28675

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-2533 <SCH>

A;Cross-references: EMBL:X96400; PIDN:CAA65264.1

C;Genetics:

A;Gene: alpha-51D

A;Genetic code: SGC5

A;Introns: 280/3; 538/2; 1248/2

C;Superfamily: G surface protein

Query Match 72.9%; Score 35; DB 2; Length 2533;  
Best Local Similarity 62.5%; Pred. No. 2.9e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
 |||: |:|  
Db 2450 CVLQSGKC 2457

Search completed: November 13, 2003, 09:52:54

Job time : 10.3333 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:31:40 ; Search time 4.58333 Seconds  
 (without alignments)  
 82.083 Million cell updates/sec

Title: US-09-228-866-4  
Perfect score: 48  
Sequence: 1 CVLRGGRC 8

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues



Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	39	81.2	3695	1	LMA5_HUMAN	O15230 homo sapien
2	38	79.2	67	1	D103_HUMAN	P81534 homo sapien
3	37	77.1	34	1	BUTH_ANDAU	P56685 androctonus
4	36	75.0	63	1	BD03_MOUSE	Q9wt10 mus musculu
5	36	75.0	65	1	PAFP_PHYAM	P81418 phytolacca
6	36	75.0	224	1	PLF1_MOUSE	P04095 mus musculu
7	36	75.0	224	1	PLF2_MOUSE	P04768 mus musculu
8	36	75.0	224	1	PLF3_MOUSE	P18918 mus musculu
9	36	75.0	437	1	PROP_MOUSE	P11680 mus musculu
10	36	75.0	522	1	IL9R_HUMAN	Q01113 homo sapien
11	36	75.0	734	1	ZN42_HUMAN	P28698 homo sapien
12	35	72.9	370	1	YR51_CAEEL	Q09321 caenorhabdi
13	35	72.9	1121	1	DPOL_ADEG1	Q64751 avian adeno
14	34	70.8	61	1	AMP1_MIRJA	P25403 mirabilis j
15	34	70.8	63	1	AMP2_MIRJA	P25404 mirabilis j
16	34	70.8	119	1	CHA1_ANTPO	P02846 antheraea p
17	34	70.8	347	1	NODI_RHIGA	P50332 rhizobium g
18	33	68.8	162	1	YLR2_EBV	P03197 epstein-bar
19	33	68.8	416	1	PROA_LEPIN	P94872 leptospira
20	33	68.8	518	1	KR2_HSV11	P04290 herpes simp
21	33	68.8	632	1	ETFD_SCHPO	P87111 s probable
22	33	68.8	697	1	YHF0_YEAST	P38721 saccharomyc
23	33	68.8	1017	1	DPOL_ADEB2	O72539 bovine aden
24	33	68.8	1023	1	DPOL_ADEB3	O72540 bovine aden
25	33	68.8	1056	1	DPOL_ADE02	P03261 human adeno
26	33	68.8	1056	1	DPOL_ADE05	P04495 human adeno
27	33	68.8	1061	1	DPOL_ADE12	P06538 human adeno
28	33	68.8	1115	1	GPCR_LYMST	P46023 lymnaea sta
29	33	68.8	1122	1	DPOL_ADE07	P05664 human adeno
30	33	68.8	1149	1	DPOL_ADECC	Q65946 canine aden
31	33	68.8	1150	1	DPOL_ADECT	P87553 canine aden
32	33	68.8	1188	1	DPOL_ADE40	P48311 human adeno
33	33	68.8	1193	1	DPOL_ADE04	P87503 human adeno
34	33	68.8	1758	1	YIR7_YEAST	P40434 saccharomyc
35	33	68.8	1790	1	LMB1_DROME	P11046 drosophila
36	33	68.8	3718	1	LMA5_MOUSE	Q61001 mus musculu
37	32	66.7	68	1	BD01_HUMAN	Q09753 homo sapien

38	32	66.7	68	1	BD01_MACMU	O18794 macaca mula
39	32	66.7	101	1	VP15_BPAPS	Q9t1t3 bacterioph
40	32	66.7	153	1	YN8K_YEAST	P53732 saccharomyc
41	32	66.7	156	1	RISB_VIBCH	Q9kpu4 vibrio chol
42	32	66.7	163	1	RISB_CORAM	O24753 corynebacte
43	32	66.7	171	1	CRIO_MOUSE	P51865 mus musculu
44	32	66.7	188	1	CRI1_HUMAN	P13385 homo sapien
45	32	66.7	188	1	CRI2_HUMAN	P51864 homo sapien

# ALIGNMENTS

## RESULT 1

### LMA5\_HUMAN

ID LMA5\_HUMAN STANDARD; PRT; 3695 AA.  
AC O15230; Q8WZA7; Q9H1P1;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Laminin alpha-5 chain precursor.  
GN LAMA5 OR KIAA0533 OR KIAA1907.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21638749; PubMed=11780052;  
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,  
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,  
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,  
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,  
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,  
RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,  
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
RA Lehvaeslaiho M.H., Liversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
RA Marsh V.L., Martin S.L., McConnell L.J., McLay K., McMurray A.A.,  
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,  
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,  
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,  
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,  
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,  
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
RA Rogers J.;  
RT "The DNA sequence and comparative analysis of human chromosome 20.";  
RL Nature 414:865-871(2001).  
RN [2]

RP SEQUENCE OF 197-1934 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=21456161; PubMed=11572484;  
 RA Nagase T., Kikuno R., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XXI.  
 RT The complete sequences of 60 new cDNA clones from brain which code for  
 RT large proteins.";  
 RL DNA Res. 8:179-187(2001).  
 RN [3]  
 RP SEQUENCE OF 2051-3695 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=98290545; PubMed=9628581;  
 RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,  
 RA Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. IX.  
 RT The complete sequences of 100 new cDNA clones from brain which can  
 RT code for large proteins in vitro.";  
 RL DNA Res. 5:31-39(1998).  
 RN [4]  
 RP SEQUENCE OF 2743-3695 FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=97415425; PubMed=9271224;  
 RA Durkin M.E., Loechel F., Mattei M.-G., Gilpin B.J., Albrechtsen R.,  
 RA Wewer U.M.;  
 RT "Tissue-specific expression of the human laminin alpha5-chain, and  
 RT mapping of the gene to human chromosome 20q13.2-13.3 and to distal  
 RT mouse chromosome 2 near the locus for the ragged (Ra) mutation.";  
 RL FEBS Lett. 411:296-300(1997).  
 RN [5]  
 RP EXPRESSION IN RETINA.  
 RX MEDLINE=20422761; PubMed=10964957;  
 RA Libby R.T., Champlaud M.-F., Claudepierre T., Xu Y., Gibbons E.P.,  
 RA Koch M., Burgeson R.E., Hunter D.D., Brunken W.J.;  
 RT "Laminin expression in adult and developing retinae: evidence of two  
 RT novel CNS laminins.";  
 RL J. Neurosci. 20:6517-6528(2000).  
 CC -!- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ  
 CC IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION AND ORGANIZATION  
 CC OF CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING  
 CC WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.  
 CC -!- SUBUNIT: LAMININ-15 COMPLEX IS AN HETEROTRIMER COMPOSED OF THREE  
 CC CHAINS (ALPHA-5/BETA-2/GAMMA-3) WHICH ARE BOUND TO EACH OTHER BY  
 CC DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE COMPRISING ONE LONG  
 CC AND THREE SHORT ARMS WITH GLOBULES AT EACH END.  
 CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT  
 CC MEMBRANES (MAJOR COMPONENT).  
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN HEART, LUNG, KIDNEY, SKELETAL  
 CC MUSCLE, PANCREAS, RETINA AND PLACENTA. LITTLE OR NO EXPRESSION IN  
 CC BRAIN AND LIVER.  
 CC -!- DOMAIN: DOMAIN G IS GLOBULAR AND IS PART OF THE MAJOR CELL-BINDING  
 CC SITE LOCATED IN THE LONG ARM OF THE LAMININ HETEROTRIMER.  
 CC -!- SIMILARITY: Contains 1 laminin N-terminal domain.  
 CC -!- SIMILARITY: Contains 22 laminin EGF-like domains.  
 CC -!- SIMILARITY: Contains 2 laminin IV domains.  
 CC -!- SIMILARITY: Contains 5 laminin G-like domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AL354836; CAC22309.1; ALT\_SEQ.  
 DR EMBL; AL354836; CAC22310.1; -.  
 DR EMBL; AB067494; BAB67800.1; -.  
 DR EMBL; AB011105; BAA25459.1; -.  
 DR EMBL; Z95636; CAB09137.1; -.  
 DR HSSP; P02468; 1KLO.  
 DR Genew; HGNC:6485; LAMA5.  
 DR MIM; 601033; -.  
 DR InterPro; IPR006209; EGF\_like.  
 DR InterPro; IPR000034; Laminin\_B.  
 DR InterPro; IPR002049; Laminin\_EGF.  
 DR InterPro; IPR001791; Laminin\_G.  
 DR InterPro; IPR001886; LamNT.  
 DR Pfam; PF00052; laminin\_B; 1.  
 DR Pfam; PF00053; laminin\_EGF; 18.  
 DR Pfam; PF00054; laminin\_G; 2.  
 DR Pfam; PF00055; laminin\_Nterm; 1.  
 DR PRINTS; PR00011; EGFLAMININ.  
 DR ProDom; PD002082; Lam\_N2; 1.  
 DR ProDom; PD003031; Laminin\_B; 1.  
 DR SMART; SM00180; EGF\_Lam; 20.  
 DR SMART; SM00281; LamB; 1.  
 DR SMART; SM00282; LamG; 5.  
 DR SMART; SM00136; LamNT; 1.  
 DR PROSITE; PS00022; EGF\_1; 19.  
 DR PROSITE; PS01186; EGF\_2; 3.  
 DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; 19.  
 DR PROSITE; PS50025; LAM\_G\_DOMAIN; 5.  
 KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;  
 KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.  
 FT SIGNAL 1 35 POTENTIAL.  
 FT CHAIN 36 3695 LAMININ ALPHA-5 CHAIN.  
 FT DOMAIN 36 299 LAMININ N-TERMINAL (DOMAIN VI).  
 FT DOMAIN 300 358 LAMININ EGF-LIKE 1.  
 FT DOMAIN 359 428 LAMININ EGF-LIKE 2.  
 FT DOMAIN 429 474 LAMININ EGF-LIKE 3.  
 FT DOMAIN 494 540 LAMININ EGF-LIKE 4.  
 FT DOMAIN 541 586 LAMININ EGF-LIKE 5.  
 FT DOMAIN 587 631 LAMININ EGF-LIKE 6.  
 FT DOMAIN 632 676 LAMININ EGF-LIKE 7.  
 FT DOMAIN 677 722 LAMININ EGF-LIKE 8.  
 FT DOMAIN 723 775 LAMININ EGF-LIKE 9.  
 FT DOMAIN 776 828 LAMININ EGF-LIKE 10.  
 FT DOMAIN 829 850 LAMININ EGF-LIKE 11 (INCOMPLETE).  
 FT DOMAIN 851 1437 LAMININ DOMAIN IV 1 (DOMAIN IV B).  
 FT DOMAIN 1438 1483 LAMININ EGF-LIKE 12.  
 FT DOMAIN 1484 1527 LAMININ EGF-LIKE 13.  
 FT DOMAIN 1528 1576 LAMININ EGF-LIKE 14.  
 FT DOMAIN 1577 1627 LAMININ EGF-LIKE 15.  
 FT DOMAIN 1628 1637 LAMININ EGF-LIKE 16 (N-TERMINAL).

FT	DOMAIN	1638	1830	LAMININ DOMAIN IV 2 (DOMAIN IV A).
FT	DOMAIN	1831	1863	LAMININ EGF-LIKE 16 (C-TERMINAL).
FT	DOMAIN	1864	1912	LAMININ EGF-LIKE 17.
FT	DOMAIN	1913	1968	LAMININ EGF-LIKE 18.
FT	DOMAIN	1969	2022	LAMININ EGF-LIKE 19.
FT	DOMAIN	2023	2069	LAMININ EGF-LIKE 20.
FT	DOMAIN	2070	2116	LAMININ EGF-LIKE 21.
FT	DOMAIN	2117	2166	LAMININ EGF-LIKE 22.
FT	DOMAIN	2167	2735	DOMAIN II AND I.
FT	DOMAIN	2736	2929	LAMININ G-LIKE 1.
FT	DOMAIN	2941	3115	LAMININ G-LIKE 2.
FT	DOMAIN	3124	3292	LAMININ G-LIKE 3.
FT	DOMAIN	3340	3513	LAMININ G-LIKE 4.
FT	DOMAIN	3520	3692	LAMININ G-LIKE 5.
FT	DOMAIN	2203	2221	COILED COIL (POTENTIAL).
FT	DOMAIN	2335	2466	COILED COIL (POTENTIAL).
FT	DOMAIN	2510	2670	COILED COIL (POTENTIAL).
FT	SITE	1722	1724	CELL ATTACHMENT SITE (POTENTIAL).
FT	SITE	1838	1840	CELL ATTACHMENT SITE (POTENTIAL).
FT	DISULFID	300	309	BY SIMILARITY.
FT	DISULFID	302	322	BY SIMILARITY.
FT	DISULFID	324	333	BY SIMILARITY.
FT	DISULFID	336	356	BY SIMILARITY.
FT	DISULFID	359	368	BY SIMILARITY.
FT	DISULFID	361	393	BY SIMILARITY.
FT	DISULFID	396	405	BY SIMILARITY.
FT	DISULFID	408	426	BY SIMILARITY.
FT	DISULFID	429	440	BY SIMILARITY.
FT	DISULFID	431	447	BY SIMILARITY.
FT	DISULFID	449	458	BY SIMILARITY.
FT	DISULFID	461	471	BY SIMILARITY.
FT	DISULFID	494	506	BY SIMILARITY.
FT	DISULFID	496	515	BY SIMILARITY.
FT	DISULFID	517	526	BY SIMILARITY.
FT	DISULFID	529	538	BY SIMILARITY.
FT	DISULFID	541	553	BY SIMILARITY.
FT	DISULFID	543	560	BY SIMILARITY.
FT	DISULFID	562	571	BY SIMILARITY.
FT	DISULFID	574	584	BY SIMILARITY.
FT	DISULFID	587	599	BY SIMILARITY.
FT	DISULFID	589	605	BY SIMILARITY.
FT	DISULFID	607	616	BY SIMILARITY.
FT	DISULFID	619	629	BY SIMILARITY.
FT	DISULFID	632	644	BY SIMILARITY.

Query Match 81.2%; Score 39; DB 1; Length 3695;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVLRGGR 7  
 |||||  
 Db 1928 CVLRGGR 1934

RESULT 2  
 D103\_HUMAN  
 ID D103\_HUMAN STANDARD; PRT; 67 AA.

AC P81534; Q9NPF6;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Beta-defensin 3 precursor (BD-3) (hBD-3) (Beta-defensin 103) (Defensin  
 DE like protein).  
 GN DEFB103 OR DEFB3 OR BD3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., SEQUENCE OF 23-67, FUNCTION, TISSUE SPECIFICITY,  
 RP INDUCTION, AND MASS SPECTROMETRY.  
 RC TISSUE=Keratinocytes, Lung epithelial cells, and Tracheal epithelium;  
 RX MEDLINE=21101950; PubMed=11085990;  
 RA Harder J., Bartels J., Christophers E., Schroeder J.-M.;  
 RT "Isolation and characterization of human deta-defensin-3, a novel  
 RT human inducible peptide antibiotic.";  
 RL J. Biol. Chem. 276:5707-5713(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RX MEDLINE=21558153; PubMed=11702237;  
 RA Garcia J.-R., Jaumann F., Schulz S., Krause A., Rodriguez-Jimenez J.,  
 RA Forssmann U., Adermann K., Kluver E., Vogelmeier C., Becker D.,  
 RA Hedrich R., Forssmann W.-G., Bals R.;  
 RT "Identification of a novel, multifunctional beta-defensin (human  
 RT beta-defensin 3) with specific antimicrobial activity. Its  
 RT interaction with plasma membranes of Xenopus oocytes and the  
 RT induction of macrophage chemoattraction.";  
 RL Cell Tissue Res. 306:257-264(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21125233; PubMed=11223260;  
 RA Jia H.P., Schutte B.C., Schudy A., Linzmeier R., Guthmiller J.M.,  
 RA Johnson G.K., Tack B.F., Mitros J.P., Rosenthal A., Ganz T.,  
 RA McCray P.B. Jr.;  
 RT "Discovery of new human defensins using a genomics-based approach.";  
 RL Gene 263:211-218(2001).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Imai Y.;  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Adler D.A., Diamond G., Sheppard P., Holloway J., Presnell S.,  
 RA Jaspers S., Whitmore T., Fox B., Gosink J., Rixon M., Gao Z.,  
 RA Haldeman B., O'Hara P.;  
 RT "EST and genomic database mining yield novel human and mouse  
 RT beta-defensins.";  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.  
 CC -!- FUNCTION: EXHIBITS ANTIMICROBIAL ACTIVITY AGAINST GRAM-POSITIVE  
 CC BACTERIA S.AUREUS AND S.PYOGENES, GRAM-NEGATIVE BACTERIA  
 CC P.AERUGINOSA AND E.COLI AND THE YEAST C.ALBICANS. KILLS  
 CC MULTIRESISTANT S.AUREUS AND VANCOMYCIN-RESISTENT E.FAECIUM. NO  
 CC SIGNIFICANT HEMOLYTIC ACTIVITY WAS OBSERVED.  
 CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN SKIN AND TONSILS, AND TO A  
CC LESSER EXTENT IN TRACHEA, UTERUS, KIDNEY, THYMUS, ADENOID, PHARYNX  
CC AND TONGUE. LOW EXPRESSION IN SALIVARY GLAND, BONE MARROW, COLON,  
CC STOMACH, POLYP AND LARYNX. NO EXPRESSION IN SMALL INTESTINE.  
CC -!- INDUCTION: BY INFECTION OF BACTERIA AND BY INTERFERON GAMMA.  
CC -!- MASS SPECTROMETRY: MW=5154.59; METHOD=Electrospray; RANGE=23-67.  
CC -!- SIMILARITY: BELONGS TO THE BETA-DEFENSIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AJ237673; CAC03097.1; -.  
DR EMBL; AF295370; AAG02237.1; -.  
DR EMBL; AF217245; AAF73853.1; -.  
DR EMBL; AB037972; BAB40572.1; -.  
DR EMBL; AF301470; AAG22030.1; -.  
DR PDB; 1KJ6; 20-MAR-02.  
DR Genew; HGNC:15967; DEFB103.  
DR MIM; 606611; -.  
DR GO; GO:0005576; C:extracellular; NAS.  
DR GO; GO:0008224; F:Gram-positive antibacterial peptide activity; TAS.  
DR GO; GO:0006965; P:anti-Gram-positive bacterial polypeptide in. . .; TAS.  
DR InterPro; IPR001855; Defensin\_beta.  
DR Pfam; PF00711; Defensin\_beta; 1.  
KW Antibiotic; Signal; 3D-structure.  
FT SIGNAL 1 22  
FT CHAIN 23 67 BETA-DEFENSIN 3.  
FT DISULFID 33 62 BY SIMILARITY.  
FT DISULFID 40 55 BY SIMILARITY.  
FT DISULFID 45 63 BY SIMILARITY.  
SQ SEQUENCE 67 AA; 7697 MW; 54266DE1C90D4B65 CRC64;

Query Match 79.2%; Score 38; DB 1; Length 67;  
Best Local Similarity 75.0%; Pred. No. 0.84;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVLRGGRC 8  
| :| | | |  
Db 33 CRVRGGRC 40

# RESULT 3

BUTH\_ANDAU

ID BUTH\_ANDAU STANDARD; PRT; 34 AA.  
AC P56685; P81617;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Buthinin.  
OS Androctonus australis hector (Sahara scorpion).  
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;  
OC Buthoidea; Buthidae; Androctonus.

OX NCBI\_TaxID=70175;  
 RN [1]  
 RP SEQUENCE, AND CHARACTERIZATION.  
 RC TISSUE=Hemolymph;  
 RX MEDLINE=97094646; PubMed=8939880;  
 RA Ehret-Sabatier L., Loew D., Goyffon M., Fehlbaum P., Hoffmann J.A.,  
 RA van Dorsselaer A., Bulet P.;  
 RT "Characterization of novel cysteine-rich antimicrobial peptides from  
 RT scorpion blood.";  
 RL J. Biol. Chem. 271:29537-29544(1996).  
 CC -!- FUNCTION: ACTIVE AGAINST BOTH GRAM-POSITIVE AND GRAM-NEGATIVE  
 CC BACTERIA.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- PTM: CONTAINS THREE DISULFIDE BONDS.  
 CC -!- MASS SPECTROMETRY: MW=3968.5; METHOD=Electrospray.  
 KW Antibiotic.  
 SQ SEQUENCE 34 AA; 3975 MW; 03323E99B7388B07 CRC64;

Query Match 77.1%; Score 37; DB 1; Length 34;  
 Best Local Similarity 75.0%; Pred. No. 0.67;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
 | |||||  
 Db 17 CGFRGGRC 24

#### RESULT 4

##### BD03\_MOUSE

ID BD03\_MOUSE STANDARD; PRT; 63 AA.  
 AC Q9WTL0;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Beta-defensin 3 precursor (BD-3) (mBD-3).  
 GN DEFB3 OR BD3.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A., FUNCTION, INDUCTION, AND TISSUE SPECIFICITY.  
 RC STRAIN=C57BL/6; TISSUE=Lung;  
 RX MEDLINE=99307216; PubMed=10377137;  
 RA Bals R., Wang X., Meegalla R.L., Wattler S., Weiner D.J., Nehls M.C.,  
 RA Wilson J.M.;  
 RT "Mouse beta-defensin 3 is an inducible antimicrobial peptide expressed  
 RT in the epithelia of multiple organs.";  
 RL Infect. Immun. 67:3542-3547(1999).  
 RN [2]  
 RP TISSUE SPECIFICITY.  
 RC STRAIN=C57BL/6, 129/SvJ, and FVB; TISSUE=Lung;  
 RX MEDLINE=20517883; PubMed=10922379;  
 RA Jia H.P., Wowk S.A., Schutte B.C., Lee S.K., Vivado A., Tack B.F.,  
 RA Bevins C.L., McCray P.B. Jr.;  
 RT "A novel murine beta-defensin expressed in tongue, esophagus, and  
 RT trachea.";



RL J. Biol. Chem. 275:33314-33320(2000).  
 CC -!- FUNCTION: ANTIMICROBIAL PEPTIDE AGAINST GRAM-NEGATIVE BACTERIA  
 CC E.COLI AND P.AERUGINOSA.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN SALIVARY GLANDS,  
 CC EPIDIDYMIS, OVARY AND PANCREAS AND TO A LESSER EXTENT IN LUNG,  
 CC LIVER AND BRAIN. LOW OR NO EXPRESSION IN SKELETAL MUSCLE AND  
 CC TONGUE.  
 CC -!- INDUCTION: By bacterial infection.  
 CC -!- SIMILARITY: BELONGS TO THE BETA-DEFENSIN FAMILY. LAP/TAP  
 CC SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF093245; AAD29573.1; -.  
 DR EMBL; AF092929; AAD29572.1; -.  
 DR HSSP; P46170; 1BNB.  
 DR MGD; MGI:1351612; Defb3.  
 DR InterPro; IPR001855; Defensin\_beta.  
 DR InterPro; IPR006080; Defensin\_mammal.  
 DR Pfam; PF00711; Defensin\_beta; 1.  
 DR SMART; SM00048; DEFSN; 1.  
 KW Antibiotic; Cleavage on pair of basic residues; Signal.  
 FT SIGNAL 1 20 POTENTIAL.  
 FT PROPEP 21 22 POTENTIAL.  
 FT CHAIN 23 63 BETA-DEFENSIN 3.  
 FT DISULFID 31 59 BY SIMILARITY.  
 FT DISULFID 38 52 BY SIMILARITY.  
 FT DISULFID 42 60 BY SIMILARITY.  
 SQ SEQUENCE 63 AA; 7126 MW; 9D59BC8AD16EA330 CRC64;

Query Match 75.0%; Score 36; DB 1; Length 63;  
 Best Local Similarity 62.5%; Pred. No. 1.9;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
 | : : : :  
 Db 31 CLRKGGRC 38

# RESULT 5

## PAFP\_PHYAM

ID PAFP\_PHYAM STANDARD; PRT; 65 AA.  
 AC P81418; O82728;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Anti-fungal protein 1 precursor (PAFP-S).  
 GN AFPS-1.  
 OS Phytolacca americana (Common pokeberry) (Virginian pokeweed).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Caryophyllidae; Caryophyllales; Phytolaccaceae; Phytolacca.  
 OX NCBI\_TaxID=3527;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Seed;  
 RA Liu Y., Ren F., Xu C., Zhao J.;  
 RT "The sequence of a cDNA encoding anti-fungal protein in Phytolacca  
 RT americana.";  
 RL Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Seed;  
 RA Liu Y., Wu G., Zhao J.;  
 RT "Chromosomal sequence of a gene encoding anti-fungal protein in  
 RT Phytolacca americana.";  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE OF 28-65.  
 RC TISSUE=Seed;  
 RA Feng S.;  
 RL Submitted (JUN-1998) to the SWISS-PROT data bank.  
 CC -!- FUNCTION: POSSESSES ANTIFUNGAL ACTIVITY.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: FOUND ONLY IN SEEDS.  
 CC -!- SIMILARITY: BELONGS TO THE AMP FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF048745; AAC05129.1; -.  
 DR EMBL; AF105062; AAD17942.1; -.  
 DR PDB; 1DKC; 13-DEC-00.  
 KW Plant defense; Fungicide; Signal; 3D-structure.  
 FT SIGNAL 1 27  
 FT CHAIN 28 65 ANTI-FUNGAL PROTEIN 1.  
 SQ SEQUENCE 65 AA; 6804 MW; 0073DE3ABBDC5B5C CRC64;

Query Match 75.0%; Score 36; DB 1; Length 65;  
 Best Local Similarity 62.5%; Pred. No. 2;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CVLRGGRC 8  
 | : | | |  
 Db 30 CIKNGGRC 37

RESULT 6  
 PLF1\_MOUSE  
 ID PLF1\_MOUSE STANDARD; PRT; 224 AA.  
 AC P04095;  
 DT 01-NOV-1986 (Rel. 03, Created)

DT 01-NOV-1986 (Rel. 03, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Proliferin 1 precursor (Mitogen-regulated protein 1).  
 GN PLF OR PLF1 OR MRP1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/c;  
 RX MEDLINE=84272617; PubMed=6087314;  
 RA Linzer D.I.H., Nathans D.;  
 RT "Nucleotide sequence of a growth-related mRNA encoding a member of  
 RT the prolactin-growth hormone family.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 81:4255-4259(1984).  
 RN [2]  
 RP SEQUENCE OF 1-10 FROM N.A.  
 RC STRAIN=BALB/c;  
 RX MEDLINE=88029317; PubMed=3478191;  
 RA Linzer D.I.H., Mordacq J.C.;  
 RT "Transcriptional regulation of proliferin gene expression in response  
 RT to serum in transfected mouse cells.";  
 RL EMBO J. 6:2281-2288(1987).  
 CC -!- FUNCTION: MAY HAVE A ROLE IN EMBRYONIC DEVELOPMENT. IT IS  
 CC LIKELY TO PROVIDE A GROWTH STIMULUS TO TARGET CELLS IN MATERNAL  
 CC AND FETAL TISSUES DURING THE DEVELOPMENT OF THE EMBRYO AT MID-  
 CC GESTATION.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X05787; CAA29231.1; -.  
 DR EMBL; K02245; AAA39946.1; -.  
 DR EMBL; X05786; CAA29230.1; -.  
 DR PIR; A05086; A05086.  
 DR HSSP; Q28632; 1AN3.  
 DR MGD; MGI:97618; Plf.  
 DR InterPro; IPR001400; Somatotropin.  
 DR Pfam; PF00103; hormone; 1.  
 DR PRINTS; PR00836; SOMATOTROPIN.  
 DR PROSITE; PS00266; SOMATOTROPIN\_1; 1.  
 DR PROSITE; PS00338; SOMATOTROPIN\_2; 1.  
 KW Hormone; Signal; Multigene family.  
 FT SIGNAL 1 29  
 FT CHAIN 30 224 PROLIFERIN 1.  
 FT DISULFID 33 40 BY SIMILARITY.  
 FT DISULFID 87 199 BY SIMILARITY.  
 FT DISULFID 216 224 BY SIMILARITY.  
 SQ SEQUENCE 224 AA; 25367 MW; 3786F100C338374B CRC64;

Query Match 75.0%; Score 36; DB 1; Length 224;  
 Best Local Similarity 62.5%; Pred. No. 6.6;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
 | : | | |  
 Db 33 CAMRNGRC 40

# RESULT 7

## PLF2\_MOUSE

ID PLF2\_MOUSE STANDARD; PRT; 224 AA.  
 AC P04768;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Proliferin 2 precursor (Mitogen-regulated protein 2).  
 GN PLF2 OR MRP2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/c;  
 RX MEDLINE=85242683; PubMed=3859868;  
 RA Linzer D.I.H., Lee S.-J., Ogren L., Talamantes F., Nathans D.;  
 RT "Identification of proliferin mRNA and protein in mouse placenta."  
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4356-4359(1985).  
 CC -!- FUNCTION: MAY HAVE A ROLE IN EMBRYONIC DEVELOPMENT. IT IS  
 CC LIKELY TO PROVIDE A GROWTH STIMULUS TO TARGET CELLS IN MATERNAL  
 CC AND FETAL TISSUES DURING THE DEVELOPMENT OF THE EMBRYO AT MID-  
 CC GESTATION.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; K03235; AAA39945.1; -.  
 DR HSSP; Q28632; 1AN3.  
 DR MGD; MGI:1341833; Plf2.  
 DR InterPro; IPR001400; Somatotropin.  
 DR Pfam; PF00103; hormone; 1.  
 DR PRINTS; PR00836; SOMATOTROPIN.  
 DR PROSITE; PS00266; SOMATOTROPIN\_1; 1.  
 DR PROSITE; PS00338; SOMATOTROPIN\_2; 1.  
 KW Hormone; Signal; Multigene family.  
 FT SIGNAL 1 29  
 FT CHAIN 30 224 PROLIFERIN 2.  
 FT DISULFID 33 40 BY SIMILARITY.

FT DISULFID 87 199 BY SIMILARITY.  
FT DISULFID 216 224 BY SIMILARITY.  
SQ SEQUENCE 224 AA; 25312 MW; 1EB34BEA21433B82 CRC64;

Query Match 75.0%; Score 36; DB 1; Length 224;  
Best Local Similarity 62.5%; Pred. No. 6.6;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
| :| |||  
Db 33 CAMRNGRC 40

# RESULT 8

## PLF3\_MOUSE

ID PLF3\_MOUSE STANDARD; PRT; 224 AA.  
AC P18918;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Proliferin 3 precursor (Mitogen-regulated protein 3).  
GN MRPPLF3 OR PLF3 OR MRP3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CD-1; TISSUE=Fibroblast;  
RX MEDLINE=90001249; PubMed=2790033;  
RA Connor A.M., Waterhouse P., Khokha R., Denhardt D.T.;  
RT "Characterization of a mouse mitogen-regulated protein/proliferin  
RT gene and its promoter: a member of the growth hormone/prolactin gene  
RT superfamily.";  
RL Biochim. Biophys. Acta 1009:75-82(1989).  
RN [2]  
RP SEQUENCE OF 208-224 FROM N.A.  
RC STRAIN=C57BL/6J;  
RX MEDLINE=94319082; PubMed=8043949;  
RA Ko M.S., Wang X., Horton J.H., Hagen M.D., Takahashi N., Maezaki Y.,  
RA Nadeau J.H.;  
RT "Genetic mapping of 40 cDNA clones on the mouse genome by PCR.";  
RL Mamm. Genome 5:349-355(1994).  
CC -!- FUNCTION: MAY HAVE A ROLE IN EMBRYONIC DEVELOPMENT. IT IS  
CC LIKELY TO PROVIDE A GROWTH STIMULUS TO TARGET CELLS IN MATERNAL  
CC AND FETAL TISSUES DURING THE DEVELOPMENT OF THE EMBRYO AT MID-  
CC GESTATION.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC -----
DR EMBL; X16009; CAA34146.1; -.
DR EMBL; X16010; CAA34146.1; JOINED.
DR EMBL; X16011; CAA34146.1; JOINED.
DR EMBL; X16012; CAA34146.1; JOINED.
DR EMBL; X16013; CAA34146.1; JOINED.
DR EMBL; U05747; AAB60482.1; -.
DR PIR; S05648; S05648.
DR HSSP; Q28632; 1AN3.
DR MGD; MGI:1347041; Mrpplf3.
DR InterPro; IPR001400; Somatotropin.
DR Pfam; PF00103; hormone; 1.
DR PROSITE; PS00266; SOMATOTROPIN_1; 1.
DR PROSITE; PS00338; SOMATOTROPIN_2; 1.
KW Hormone; Signal; Multigene family.
FT SIGNAL      1      29
FT CHAIN       30     224      PROLIFERIN 3.
FT DISULFID    33      40      BY SIMILARITY.
FT DISULFID    87     199      BY SIMILARITY.
FT DISULFID   216     224      BY SIMILARITY.
SQ SEQUENCE    224 AA;  25338 MW;  C87F3A2310C91320 CRC64;

```

```

Query Match          75.0%;  Score 36;  DB 1;  Length 224;
Best Local Similarity 62.5%;  Pred. No. 6.6;
Matches      5;  Conservative      1;  Mismatches      2;  Indels      0;  Gaps      0;

```

```

Qy      1 CVLRGGRC 8
        | :| |||
Db      33 CAMRNGRC 40

```

# RESULT 9

## PROP\_MOUSE

```

ID  PROP_MOUSE      STANDARD;      PRT;      437 AA.
AC  P11680;
DT  01-AUG-1992 (Rel. 23, Created)
DT  01-AUG-1992 (Rel. 23, Last sequence update)
DT  28-FEB-2003 (Rel. 41, Last annotation update)
DE  Properdin (Factor P) (Fragment).
GN  PFC.
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX  NCBI_TaxID=10090;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=Macrophage;
RX  MEDLINE=88318954; PubMed=3045564;
RA  Goundis A., Reid K.B.M.;
RT  "Properdin, the terminal complement components, thrombospondin and
RT  the circumsporozoite protein of malaria parasites contain similar
RT  sequence motifs.";
RL  Nature 335:82-85(1988).
CC  -!- FUNCTION: A POSITIVE REGULATOR OF THE ALTERNATE PATHWAY OF
CC      COMPLEMENT. IT BINDS TO AND STABILIZES THE C3-AND C5-CONVERTASE
CC      ENZYME COMPLEXES.
CC  -!- SIMILARITY: Contains 6 TSP type-1 domains.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X12905; CAA31389.1; -.
DR PIR; S05478; S05478.
DR MGD; MGI:97545; Pfc.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; tsp_1; 6.
DR SMART; SM00209; TSP1; 6.
DR PROSITE; PS50092; TSP1; 6.
KW Complement alternate pathway; Glycoprotein; Repeat.
FT NON_TER      1      1
FT DOMAIN        46     103      TSP TYPE-1 1.
FT DOMAIN       105     160      TSP TYPE-1 2.
FT DOMAIN       162     224      TSP TYPE-1 3.
FT DOMAIN       226     282      TSP TYPE-1 4.
FT DOMAIN       284     345      TSP TYPE-1 5.
FT DOMAIN       347     430      TSP TYPE-1 6.
FT CARBOHYD      396     396      N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE     437 AA; 47538 MW; 2B8DBCE22B3B78BE CRC64;

```

```

Query Match          75.0%; Score 36; DB 1; Length 437;
Best Local Similarity 75.0%; Pred. No. 13;
Matches      6; Conservative      1; Mismatches      1; Indels      0; Gaps      0;

```

```

Qy      1 CVLRGGRC 8
        || |||:|
Db      73 CVGRGGQC 80

```

```

RESULT 10
IL9R_HUMAN
ID IL9R_HUMAN STANDARD; PRT; 522 AA.
AC Q01113; Q14634; Q8WWU1; Q96TF0;
DT 01-APR-1993 (Rel. 25, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Interleukin-9 receptor precursor (IL-9R).
GN (IL9RX OR IL9R) AND (IL9RY OR IL9R).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92302307; PubMed=1376929;
RA Renauld J.C., Druetz C., Kermouni A., Houssiau F., Uyttenhove C.,
RA van Roost E., van Snick J.;
RT "Expression cloning of the murine and human interleukin 9 receptor
RT cDNAs.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:5690-5694(1992).

```

RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94250901; PubMed=8193355;  
 RA Chang M.S., Engel G., Benedict C., Basu R., McNinch J.;  
 RT "Isolation and characterization of the human interleukin-9 receptor  
 RT gene.";  
 RL Blood 83:3199-3205(1994).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Melanoma;  
 RX MEDLINE=96115587; PubMed=8666384;  
 RA Kermouni A., van Roost E., Arden K.C., Vermeesch J.R., Weiss S.,  
 RA Godelaine D., Flint J., Lurquin C., Szikora J.P., Higgs D.R.,  
 RA Marynen P., Renauld J.C.;  
 RT "The IL-9 receptor gene (IL9R): genomic structure, chromosomal  
 RT localization in the pseudoautosomal region of the long arm of the sex  
 RT chromosomes, and identification of IL9R pseudogenes at 9qter, 10pter,  
 RT 16pter, and 18pter.";  
 RL Genomics 29:371-382(1995).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20122249; PubMed=10655549;  
 RA Ciccodicola A., D'Esposito M., Esposito T., Gianfrancesco F.,  
 RA Migliaccio C., Miano M.G., Matarazzo M.R., Vacca M., Franze A.,  
 RA Cuccurese M., Cocchia M., Curci A., Terracciano A., Torino A.,  
 RA Cocchia S., Mercadante G., Pannone E., Archidiacono N., Rocchi M.,  
 RA Schlessinger D., D'Urso M.;  
 RT "Differentially regulated and evolved genes in the fully sequenced  
 RT Xq/Yq pseudoautosomal region.";  
 RL Hum. Mol. Genet. 9:395-401(2000).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Rieder M.J., Armel T.Z., Carrington D.P., Chung M.-W., Lee K.L.,  
 RA Poel C.L., Toth E.J., Yi Q., Nickerson D.A.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.  
 CC -!- FUNCTION: THIS IS A RECEPTOR FOR INTERLEUKIN-9.  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein and secreted.  
 CC -!- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.  
 CC -!- SIMILARITY: Contains 1 fibronectin type III domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M84747; AAA58679.1; -.  
 DR EMBL; S71404; AAB30844.2; ALT\_SEQ.  
 DR EMBL; S71420; AAD14081.1; -.  
 DR EMBL; L39064; AAC29513.1; -.  
 DR EMBL; AJ271736; CAB96817.1; -.  
 DR EMBL; AY071830; AAL55435.1; -.  
 DR PIR; B45268; B45268.  
 DR Genew; HGNC:6030; IL9R.  
 DR MIM; 300007; -.



DR GO; GO:0005615; C:extracellular space; TAS.  
 DR GO; GO:0005887; C:integral to plasma membrane; TAS.  
 DR GO; GO:0004919; F:interleukin-9 receptor activity; TAS.  
 DR GO; GO:0008283; P:cell proliferation; TAS.  
 DR GO; GO:0007165; P:signal transduction; TAS.  
 DR InterPro; IPR002996; CR1A.  
 DR InterPro; IPR003531; Hemtopoptn\_S\_F1.  
 DR PROSITE; PS01355; HEMATOPO\_REC\_S\_F1; 1.  
 KW Receptor; Transmembrane; Glycoprotein; Signal; Polymorphism.  
 FT SIGNAL 1 40 POTENTIAL.  
 FT CHAIN 41 522 INTERLEUKIN-9 RECEPTOR.  
 FT DOMAIN 41 270 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 271 291 POTENTIAL.  
 FT DOMAIN 292 522 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 150 244 FIBRONECTIN TYPE-III.  
 FT DOMAIN 429 439 POLY-SER.  
 FT DOMAIN 440 443 POLY-ASN.  
 FT CARBOHYD 117 117 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 156 156 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARIANT 239 239 E -> Q (IN dbSNP:6522).  
 FT /FTid=VAR\_014804.  
 FT CONFLICT 331 331 G -> R (IN REF. 1 AND 2).  
 FT CONFLICT 439 439 MISSING (IN REF. 3 AND 4).  
 SQ SEQUENCE 522 AA; 57233 MW; BBB73D6E2FAE37CB CRC64;

Query Match 75.0%; Score 36; DB 1; Length 522;  
 Best Local Similarity 62.5%; Pred. No. 15;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
 | : | | | |  
 Db 95 CILRGSEC 102

RESULT 11  
 ZN42\_HUMAN  
 ID ZN42\_HUMAN STANDARD; PRT; 734 AA.  
 AC P28698; Q9UBW2;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Zinc finger protein 42 (Myeloid zinc finger 1) (MZF-1).  
 GN ZNF42 OR MZF1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM MZF1A).  
 RX MEDLINE=91317761; PubMed=1860835;  
 RA Hromas R., Collins S.J., Hickstein D., Raskind W., Deaven L.L.,  
 RA O'Hara P., Hagen F.S., Kaushansky K.;  
 RT "A retinoic acid-responsive human zinc finger gene, MZF-1,  
 RT preferentially expressed in myeloid cells."  
 RL J. Biol. Chem. 266:14183-14187(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM MZF1B-C).

RC TISSUE=Bone marrow;  
 RX MEDLINE=20432092; PubMed=10974541;  
 RA Peterson M.J., Morris J.F.;  
 RT "Human myeloid zinc finger gene MZF produces multiple transcripts and  
 RT encodes a SCAN box protein.";  
 RL Gene 254:105-118(2000).  
 CC -!- FUNCTION: MAY BE ONE REGULATOR OF TRANSCRIPTIONAL EVENTS DURING  
 CC HEMOPOIETIC DEVELOPMENT.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=MZF1A;  
 CC IsoId=P28698-1; Sequence=Displayed;  
 CC Name=MZF1B-C;  
 CC IsoId=P28698-2; Sequence=VSP\_006889, VSP\_006890;  
 CC -!- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN DIFFERENTIATING  
 CC MYELOID CELLS.  
 CC -!- INDUCTION: By retinoic acid.  
 CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
 CC FINGER PROTEINS.  
 CC -!- SIMILARITY: Contains 1 SCAN box domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M58297; AAA59898.1; -.  
 DR EMBL; AF055078; AAD55810.1; -.  
 DR EMBL; AF055077; AAD55809.1; -.  
 DR EMBL; AF161886; AAF80465.1; -.  
 DR HSSP; P08047; 1SP2.  
 DR TRANSFAC; T00529; -.  
 DR Genew; HGNC:13108; ZNF42.  
 DR MIM; 194550; -.  
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; TAS.  
 DR InterPro; IPR003309; Treg\_SCAN.  
 DR InterPro; IPR007087; Znf\_C2H2.  
 DR InterPro; IPR007086; Znf\_C2H2\_sub.  
 DR Pfam; PF02023; SCAN; 1.  
 DR Pfam; PF00096; zf-C2H2; 13.  
 DR PRINTS; PR00048; ZINCFINGER.  
 DR ProDom; PD000003; Znf\_C2H2; 6.  
 DR SMART; SM00431; LER; 1.  
 DR SMART; SM00355; Znf\_C2H2; 13.  
 DR PROSITE; PS50804; SCAN\_BOX; 1.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 13.  
 DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 13.  
 KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;  
 KW Nuclear protein; Repeat; Alternative splicing; Polymorphism.  
 FT DOMAIN 44 125 SCAN BOX.  
 FT DOMAIN 310 321 ASP/GLU-RICH (ACIDIC).  
 FT ZN\_FING 356 378 C2H2-TYPE.  
 FT ZN\_FING 384 406 C2H2-TYPE.

FT	ZN_FING	412	434	C2H2-TYPE.
FT	ZN_FING	440	462	C2H2-TYPE.
FT	DOMAIN	463	484	GLY/PRO-RICH.
FT	ZN_FING	485	507	C2H2-TYPE.
FT	ZN_FING	513	535	C2H2-TYPE.
FT	ZN_FING	541	563	C2H2-TYPE.
FT	ZN_FING	569	591	C2H2-TYPE.
FT	ZN_FING	597	619	C2H2-TYPE.
FT	ZN_FING	625	647	C2H2-TYPE.
FT	ZN_FING	653	675	C2H2-TYPE.
FT	ZN_FING	681	703	C2H2-TYPE.
FT	ZN_FING	709	731	C2H2-TYPE.
FT	VARSPPLIC	1	249	Missing (in isoform MZF1B-C).
FT				/FTId=VSP_006889.
FT	VARSPPLIC	250	257	EAGGIFSP -> MNGPLVYA (in isoform
FT				MZF1B-C).
FT				/FTId=VSP_006890.
FT	VARIANT	331	331	I -> V (IN dbSNP:4756).
FT				/FTId=VAR_014826.
FT	CONFLICT	304	305	AL -> RV (IN REF. 1).
SQ	SEQUENCE	734 AA;	82036 MW;	2BE7D69B18F29437 CRC64;

---

Query Match 75.0%; Score 36; DB 1; Length 734;  
 Best Local Similarity 85.7%; Pred. No. 21;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VLRGGRC 8  
 |:|||||  
 Db 352 VVRGGRC 358

# RESULT 12

YR51\_CAEEL

ID YR51\_CAEEL STANDARD; PRT; 370 AA.  
 AC Q09321;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Hypothetical 42.0 kDa protein F42A8.1 in chromosome II.  
 GN F42A8.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Matthews P.;  
 RL Submitted (JAN-1995) to the EMBL/GenBank/DDBJ databases.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; Z47809; CAA87779.1; -.  
 DR PIR; T22082; T22082.  
 DR WormPep; F42A8.1; CE01578.  
 KW Hypothetical protein.  
 SQ SEQUENCE 370 AA; 41975 MW; F1C2C5C9E0956034 CRC64;

Query Match 72.9%; Score 35; DB 1; Length 370;  
 Best Local Similarity 62.5%; Pred. No. 17;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
 | | : | | : |  
 Db 209 CKLQGGKC 216

# RESULT 13

DPOL\_ADEG1

ID DPOL\_ADEG1 STANDARD; PRT; 1121 AA.

AC Q64751;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE DNA polymerase (EC 2.7.7.7).

GN POL.

OS Avian adenovirus gall (strain Phelps) (Fowl adenovirus 1) (CELO).

OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Aviadenovirus.

OX NCBI\_TaxID=10553;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=96186720; PubMed=8627769;

RA Chiocca S., Kurzbauer R., Schaffner G., Baker A., Mautner V.,

RA Cotten M.;

RT "The complete DNA sequence and genomic organization of the avian

RT adenovirus CELO.";

RL J. Virol. 70:2939-2949(1996).

CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate

CC + {DNA}(N).

CC -!- MISCELLANEOUS: THIS DNA POLYMERASE REQUIRES A PROTEIN AS A PRIMER.

CC -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; U46933; AAC54904.1; -.

DR InterPro; IPR006172; DNA\_pol\_B.

DR InterPro; IPR004868; DNA\_pol\_B\_2.

DR Pfam; PF03175; DNA\_pol\_B\_2; 1.

DR PRINTS; PR00106; DNAPOLB.

DR SMART; SM00486; POLBc; 1.

DR PROSITE; PS00116; DNA\_POLYMERASE\_B; 1.

KW Transferase; DNA-directed DNA polymerase; DNA replication;

KW DNA-binding.

SQ SEQUENCE 1121 AA; 129395 MW; A55B9B6A54D3BDE1 CRC64;

Query Match 72.9%; Score 35; DB 1; Length 1121;

Best Local Similarity 100.0%; Pred. No. 49;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 LRGGRG 8

|||||

Db 488 LRGGRG 493

#### RESULT 14

##### AMP1\_MIRJA

ID AMP1\_MIRJA STANDARD; PRT; 61 AA.

AC P25403;

DT 01-MAY-1992 (Rel. 22, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Antimicrobial peptide 1 precursor (AMP1) (MJ-AMP1) (Fragment).

GN AMP1.

OS Mirabilis jalapa (Garden four-o'clock).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Caryophyllidae; Caryophyllales; Nyctaginaceae; Mirabilis.

OX NCBI\_TaxID=3538;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Seed;

RX MEDLINE=95375234; PubMed=7647302;

RA de Bolle M.F., Eggermont K., Duncan R.E., Osborn R.W., Terras F.R.G.,

RA Broekaert W.F.;

RT "Cloning and characterization of two cDNA clones encoding seed-

RT specific antimicrobial peptides from Mirabilis jalapa L.";

RL Plant Mol. Biol. 28:713-721(1995).

RN [2]

RP SEQUENCE OF 25-61.

RC TISSUE=Seed;

RX MEDLINE=92129292; PubMed=1733929;

RA Cammue B.P.A., de Bolle M.F.C., Terras F.R.G., Proost P.,

RA van Damme J., Rees S.B., Vanderleyden J., Broekaert W.F.;

RT "Isolation and characterization of a novel class of plant

RT antimicrobial peptides from Mirabilis jalapa L. seeds.";

RL J. Biol. Chem. 267:2228-2233(1992).

CC -!- FUNCTION: POSSESSES ANTIFUNGAL ACTIVITY AND IS ALSO ACTIVE ON TWO

CC TESTED GRAM-POSITIVE BACTERIA BUT IS NONTOXIC FOR GRAM-NEGATIVE

CC BACTERIA AND CULTURED HUMAN CELLS.

CC -!- SUBUNIT: Homodimer.

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- TISSUE SPECIFICITY: FOUND ONLY IN SEEDS.

CC -!- PTM: THREE DISULFIDE BONDS ARE PRESENT.

CC -!- SIMILARITY: BELONGS TO THE AMP FAMILY.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; U15538; AAA80484.1; -.  
DR EMBL; A27777; CAA01890.1; -.  
DR PIR; S57815; S57815.  
KW Plant defense; Fungicide; Antibiotic; Signal;  
KW Pyrrolidone carboxylic acid.  
FT NON\_TER 1 1  
FT SIGNAL <1 24  
FT CHAIN 25 61 ANTIMICROBIAL PEPTIDE 1.  
FT MOD\_RES 25 25 PYRROLIDONE CARBOXYLIC ACID.  
SQ SEQUENCE 61 AA; 6605 MW; 1957BF5FC2FE75C2 CRC64;

Query Match 70.8%; Score 34; DB 1; Length 61;  
Best Local Similarity 62.5%; Pred. No. 4.5;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
|: ||||  
Db 26 CIGNGGRC 33

---

RESULT 15

AMP2\_MIRJA

ID AMP2\_MIRJA STANDARD; PRT; 63 AA.  
AC P25404;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Antimicrobial peptide 2 precursor (AMP2) (MJ-AMP2).  
GN AMP2.  
OS Mirabilis jalapa (Garden four-o'clock).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Caryophyllidae; Caryophyllales; Nyctaginaceae; Mirabilis.  
OX NCBI\_TaxID=3538;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Seed;  
RX MEDLINE=95375234; PubMed=7647302;  
RA de Bolle M.F., Eggermont K., Duncan R.E., Osborn R.W., Terras F.R.,  
RA Broekaert W.F.;  
RT "Cloning and characterization of two cDNA clones encoding seed-  
RT specific antimicrobial peptides from Mirabilis jalapa L.";  
RL Plant Mol. Biol. 28:713-721(1995).  
RN [2]  
RP SEQUENCE OF 28-63.  
RC TISSUE=Seed;  
RX MEDLINE=92129292; PubMed=1733929;  
RA Cammue B.P.A., de Bolle M.F.C., Terras F.R.G., Proost P.,  
RA van Damme J., Rees S.B., Vanderleyden J., Broekaert W.F.;  
RT "Isolation and characterization of a novel class of plant  
RT antimicrobial peptides from Mirabilis jalapa L. seeds.";  
RL J. Biol. Chem. 267:2228-2233(1992).  
CC -!- FUNCTION: POSSESSES ANTIFUNGAL ACTIVITY AND IS ALSO ACTIVE ON TWO

CC TESTED GRAM-POSITIVE BACTERIA BUT IS NONTOXIC FOR GRAM-NEGATIVE  
 CC BACTERIA AND CULTURED HUMAN CELLS.  
 CC -!- SUBUNIT: Homodimer.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: FOUND ONLY IN SEEDS.  
 CC -!- PTM: THREE DISULFIDE BONDS ARE PRESENT.  
 CC -!- SIMILARITY: BELONGS TO THE AMP FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U15539; AAA80485.1; -.  
 DR EMBL; A27779; CAA01891.1; -.  
 DR PIR; S57816; S57816.  
 KW Plant defense; Fungicide; Antibiotic; Signal.  
 FT SIGNAL 1 27  
 FT CHAIN 28 63 ANTIMICROBIAL PEPTIDE 2.  
 SQ SEQUENCE 63 AA; 6842 MW; E234721728590A84 CRC64;

Query Match 70.8%; Score 34; DB 1; Length 63;  
 Best Local Similarity 62.5%; Pred. No. 4.6;  
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
 | : | | | |  
 Db 28 CIGNGGRC 35

Search completed: November 13, 2003, 09:46:33  
 Job time : 5.58333 secs

GenCore version 5.1.6  
 Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:31:40 ; Search time 21.0833 Seconds  
 (without alignments)  
 97.917 Million cell updates/sec

Title: US-09-228-866-4  
 Perfect score: 48  
 Sequence: 1 CVLRGGRC 8

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp Vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query		DB	ID	Description
		% Match	Length			
1	41	85.4	4123	4	O75851	O75851 homo sapien
2	40	83.3	377	13	Q9IAF9	Q9iaf9 ivindomyrus
3	40	83.3	377	13	Q9IAE9	Q9iae9 mormyrus ru
4	40	83.3	377	13	Q9IAH4	Q9iah4 brienomyrus
5	40	83.3	377	13	Q9IAF0	Q9iaf0 mormyrus ov
6	40	83.3	377	13	Q9IAE8	Q9iae8 myomyrus ma
7	40	83.3	377	13	Q9IAH2	Q9iah2 brienomyrus
8	40	83.3	377	13	Q9I867	Q9i867 campylomorm
9	40	83.3	377	13	Q9IAG0	Q9iag0 isichthys h
10	40	83.3	377	13	Q9IAG1	Q9iag1 hyperopisus
11	40	83.3	377	13	Q9IAE5	Q9iae5 petrocephal
12	40	83.3	377	13	Q9IAF4	Q9iaf4 marcusenius
13	40	83.3	377	13	Q9IAD8	Q9iad8 stomatorhin
14	40	83.3	377	13	Q9IAH1	Q9iah1 brienomyrus
15	40	83.3	377	13	Q9IAE2	Q9iae2 petrocephal
16	40	83.3	377	13	Q9IAE0	Q9iae0 pollimyrus
17	40	83.3	377	13	Q9IAF6	Q9iaf6 marcusenius
18	40	83.3	377	13	Q9IAH5	Q9iah5 brienomyrus
19	40	83.3	377	13	Q9IAD9	Q9iad9 pollimyrus
20	40	83.3	377	13	Q9IAH0	Q9iah0 campylomorm
21	40	83.3	377	13	Q9IAE1	Q9iae1 pollimyrus
22	40	83.3	377	13	Q9IAG4	Q9iag4 hippopotamy



23	40	83.3	377	13	Q9IAF5	Q9iaf5 marcusenius
24	40	83.3	377	13	Q9IAE3	Q9iae3 petrocephal
25	40	83.3	377	13	Q9IAE6	Q9iae6 paramormyro
26	40	83.3	377	13	Q9IAG3	Q9iag3 hippopotamy
27	40	83.3	377	13	Q9IAD5	Q9iad5 stomatorhin
28	40	83.3	377	13	Q9IAF7	Q9iaf7 marcusenius
29	40	83.3	377	13	Q9IAF2	Q9iaf2 mormyrops n
30	40	83.3	377	13	Q9IAF1	Q9iaf1 mormyrops z
31	40	83.3	377	13	Q9IAG6	Q9iag6 gnathonemus
32	40	83.3	377	13	Q9IAD7	Q9iad7 stomatorhin
33	40	83.3	377	13	Q9IAG2	Q9iag2 hippopotamy
34	40	83.3	377	13	Q9IAD6	Q9iad6 stomatorhin
35	40	83.3	377	13	Q9IAF8	Q9iaf8 marcusenius
36	40	83.3	377	13	Q9IAE4	Q9iae4 petrocephal
37	40	83.3	377	13	Q9IAG7	Q9iag7 genyomys
38	40	83.3	377	13	Q9IAF3	Q9iaf3 mormyrops m
39	40	83.3	377	13	Q9IAG9	Q9iag9 campylomorm
40	40	83.3	377	13	Q9IAH3	Q9iah3 brienomys
41	40	83.3	377	13	Q9IAH6	Q9iah6 boulengerom
42	40	83.3	377	13	Q8AWR8	Q8awr8 pollimys
43	39	81.2	3695	4	Q8TDF8	Q8tdf8 homo sapien
44	38	79.2	64	6	Q95JD2	Q95jd2 pan troglod
45	38	79.2	67	4	Q8NFG6	Q8nfg6 homo sapien

# ALIGNMENTS

## RESULT 1

O75851

ID O75851 PRELIMINARY; PRT; 4123 AA.  
AC O75851;  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE WUGSC:H\_DJ0751H13.1 protein (Fragment).  
GN WUGSC:H\_DJ0751H13.1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Leonard S., Graves T., Strowmatt C.;  
RT "The sequence of Homo sapiens PAC clone RP4-751H13.";  
RL Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Waterston R.;  
RL Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AC004877; AAC36301.1; -.  
DR HSSP; P01130; 1AJJ.  
DR InterPro; IPR000923; BlueCu\_1.  
DR InterPro; IPR001064; Crystallin.  
DR InterPro; IPR006209; EGF\_like.  
DR InterPro; IPR000421; FA58\_C.  
DR InterPro; IPR002223; Kunitz\_BPTI.

DR InterPro; IPR002172; LDL\_receptor\_A.  
 DR InterPro; IPR002919; TIL\_Cysrich.  
 DR InterPro; IPR000884; TSP1.  
 DR InterPro; IPR001007; VWF\_C.  
 DR InterPro; IPR001846; VWF\_D.  
 DR Pfam; PF00754; F5\_F8\_type\_C; 1.  
 DR Pfam; PF00057; ldl\_recept\_a; 11.  
 DR Pfam; PF01826; TIL; 5.  
 DR Pfam; PF00090; tsp\_1; 14.  
 DR Pfam; PF00094; vwd; 3.  
 DR PRINTS; PR00261; LDLRECEPTOR.  
 DR SMART; SM00231; FA58C; 1.  
 DR SMART; SM00192; LDLa; 10.  
 DR SMART; SM00209; TSP1; 14.  
 DR SMART; SM00214; VWC; 1.  
 DR SMART; SM00216; VWD; 3.  
 DR PROSITE; PS00280; BPTI\_KUNITZ\_1; 1.  
 DR PROSITE; PS00196; COPPER\_BLUE; 1.  
 DR PROSITE; PS00225; CRYSTALLIN\_BETAGAMMA; 1.  
 DR PROSITE; PS00022; EGF\_1; 1.  
 DR PROSITE; PS01209; LDLRA\_1; 9.  
 DR PROSITE; PS50068; LDLRA\_2; 9.  
 DR PROSITE; PS50092; TSP1; 14.

FT NON\_TER 1 1  
 SQ SEQUENCE 4123 AA; 434981 MW; 7AAB6FE8DCE012FB CRC64;

Query Match 85.4%; Score 41; DB 4; Length 4123;  
 Best Local Similarity 87.5%; Pred. No. 65;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
 ||||| |  
 Db 2221 CVLRGGPC 2228

## RESULT 2

### Q9IAF9

ID Q9IAF9 PRELIMINARY; PRT; 377 AA.  
 AC Q9IAF9;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Recombination-activating protein 2 (Fragment).  
 OS Ivindomyrus opdenboschi.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;  
 OC Osteoglossiformes; Mormyridae; Ivindomyrus.  
 OX NCBI\_TaxID=91727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20115608; PubMed=10648209;  
 RA Sullivan J.P., Lavoue S., Hopkins C.D.;  
 RT "Molecular systematics of the African electric fishes (Mormyroidea:  
 RT Teleostei) and a model for the evolution of their electric organs."  
 RL J. Exp. Biol. 203:665-683(2000).  
 DR EMBL; AF201635; AAF43346.1; -.  
 DR InterPro; IPR004321; RAG2.

DR Pfam; PF03089; RAG2; 1.  
FT NON\_TER 1 1  
FT NON\_TER 377 377  
SQ SEQUENCE 377 AA; 41428 MW; B60EDE613EA0FDBE CRC64;

Query Match 83.3%; Score 40; DB 13; Length 377;  
Best Local Similarity 87.5%; Pred. No. 10;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
||| ||||  
Db 103 CVLFGGRC 110

#### RESULT 3

Q9IAE9

ID Q9IAE9 PRELIMINARY; PRT; 377 AA.

AC Q9IAE9;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE Recombination-activating protein 2 (Fragment).

OS Mormyrus rume.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;

OC Osteoglossiformes; Mormyridae; Mormyrus.

OX NCBI\_TaxID=91731;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20115608; PubMed=10648209;

RA Sullivan J.P., Lavoue S., Hopkins C.D.;

RT "Molecular systematics of the African electric fishes (Mormyroidea:

RT Teleostei) and a model for the evolution of their electric organs.";

RL J. Exp. Biol. 203:665-683(2000).

DR EMBL; AF201645; AAF43356.1; -.

DR InterPro; IPR004321; RAG2.

DR Pfam; PF03089; RAG2; 1.

FT NON\_TER 1 1

FT NON\_TER 377 377

SQ SEQUENCE 377 AA; 41364 MW; D59BAC6D739AEE56 CRC64;

Query Match 83.3%; Score 40; DB 13; Length 377;  
Best Local Similarity 87.5%; Pred. No. 10;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
||| ||||  
Db 103 CVLFGGRC 110

#### RESULT 4

Q9IAH4

ID Q9IAH4 PRELIMINARY; PRT; 377 AA.

AC Q9IAH4;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE Recombination-activating protein 2 (Fragment).  
 OS Brienomyrus hopkinsi.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;  
 OC Osteoglossiformes; Mormyridae; Brienomyrus.  
 OX NCBI\_TaxID=112141;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20115608; PubMed=10648209;  
 RA Sullivan J.P., Lavoue S., Hopkins C.D.;  
 RT "Molecular systematics of the African electric fishes (Mormyroidea:  
 RT Teleostei) and a model for the evolution of their electric organs.";  
 RL J. Exp. Biol. 203:665-683(2000).  
 DR EMBL; AF201618; AAF43329.1; -.  
 DR InterPro; IPR004321; RAG2.  
 DR Pfam; PF03089; RAG2; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 377 377  
 SQ SEQUENCE 377 AA; 41403 MW; 0A4599C6604C8123 CRC64;

Query Match 83.3%; Score 40; DB 13; Length 377;  
 Best Local Similarity 87.5%; Pred. No. 10;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
 ||| ||||  
 Db 103 CVLFGGRC 110

#### RESULT 5

##### Q9IAF0

ID Q9IAF0 PRELIMINARY; PRT; 377 AA.  
 AC Q9IAF0;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Recombination-activating protein 2 (Fragment).  
 OS Mormyrus ovis.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;  
 OC Osteoglossiformes; Mormyridae; Mormyrus.  
 OX NCBI\_TaxID=112155;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20115608; PubMed=10648209;  
 RA Sullivan J.P., Lavoue S., Hopkins C.D.;  
 RT "Molecular systematics of the African electric fishes (Mormyroidea:  
 RT Teleostei) and a model for the evolution of their electric organs.";  
 RL J. Exp. Biol. 203:665-683(2000).  
 DR EMBL; AF201644; AAF43355.1; -.  
 DR InterPro; IPR004321; RAG2.  
 DR Pfam; PF03089; RAG2; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 377 377  
 SQ SEQUENCE 377 AA; 41431 MW; 7EB7C6C644E569DB CRC64;

Query Match 83.3%; Score 40; DB 13; Length 377;

Best Local Similarity 87.5%; Pred. No. 10;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
||| |||  
Db 103 CVLFGGRC 110

RESULT 6

Q9IAE8

ID Q9IAE8 PRELIMINARY; PRT; 377 AA.  
AC Q9IAE8;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Recombination-activating protein 2 (Fragment).  
OS Myomyrus macrops.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;  
OC Osteoglossiformes; Mormyridae; Myomyrus.  
OX NCBI\_TaxID=112156;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20115608; PubMed=10648209;  
RA Sullivan J.P., Lavoue S., Hopkins C.D.;  
RT "Molecular systematics of the African electric fishes (Mormyroidea:  
RT Teleostei) and a model for the evolution of their electric organs.";  
RL J. Exp. Biol. 203:665-683(2000).  
DR EMBL; AF201646; AAF43357.1; -.  
DR InterPro; IPR004321; RAG2.  
DR Pfam; PF03089; RAG2; 1.  
FT NON\_TER 1 1  
FT NON\_TER 377 377  
SQ SEQUENCE 377 AA; 41482 MW; A8B7F60D40B6AE5E CRC64;

Query Match 83.3%; Score 40; DB 13; Length 377;  
Best Local Similarity 87.5%; Pred. No. 10;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
||| |||  
Db 103 CVLFGGRC 110

RESULT 7

Q9IAH2

ID Q9IAH2 PRELIMINARY; PRT; 377 AA.  
AC Q9IAH2;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Recombination-activating protein 2 (Fragment).  
OS Brienomyrus niger.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;  
OC Osteoglossiformes; Mormyridae; Brienomyrus.  
OX NCBI\_TaxID=42637;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20115608; PubMed=10648209;  
 RA Sullivan J.P., Lavoue S., Hopkins C.D.;  
 RT "Molecular systematics of the African electric fishes (Mormyroidea:  
 RT Teleostei) and a model for the evolution of their electric organs.";  
 RL J. Exp. Biol. 203:665-683(2000).  
 DR EMBL; AF201620; AAF43331.1; -.  
 DR InterPro; IPR004321; RAG2.  
 DR Pfam; PF03089; RAG2; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 377 377  
 SQ SEQUENCE 377 AA; 41522 MW; 2E93DC79A8B6EC4A CRC64;

Query Match 83.3%; Score 40; DB 13; Length 377;  
 Best Local Similarity 87.5%; Pred. No. 10;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
 ||| ||||  
 Db 103 CVLFGGRC 110

# RESULT 8

Q9I867

ID Q9I867 PRELIMINARY; PRT; 377 AA.  
 AC Q9I867;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Recombination-activating protein 2 (Fragment).  
 OS Campylomormyrus tamandua.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;  
 OC Osteoglossiformes; Mormyridae; Campylomormyrus.  
 OX NCBI\_TaxID=91719;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20115608; PubMed=10648209;  
 RA Sullivan J.P., Lavoue S., Hopkins C.D.;  
 RT "Molecular systematics of the African electric fishes (Mormyroidea:  
 RT Teleostei) and a model for the evolution of their electric organs.";  
 RL J. Exp. Biol. 203:665-683(2000).  
 DR EMBL; AF201625; AAF43336.1; -.  
 DR EMBL; AF201624; AAF43335.1; -.  
 DR InterPro; IPR004321; RAG2.  
 DR Pfam; PF03089; RAG2; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 377 377  
 SQ SEQUENCE 377 AA; 41387 MW; D52A9E361A56AB43 CRC64;

Query Match 83.3%; Score 40; DB 13; Length 377;  
 Best Local Similarity 87.5%; Pred. No. 10;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
 ||| ||||

Db 103 CVLFGGRC 110

RESULT 9

Q9IAG0

ID Q9IAG0 PRELIMINARY; PRT; 377 AA.  
AC Q9IAG0;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Recombination-activating protein 2 (Fragment).  
OS *Isichthys henryi*.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;  
OC Osteoglossiformes; Mormyridae; *Isichthys*.  
OX NCBI\_TaxID=112151;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20115608; PubMed=10648209;  
RA Sullivan J.P., Lavoue S., Hopkins C.D.;  
RT "Molecular systematics of the African electric fishes (Mormyroidea:  
RT Teleostei) and a model for the evolution of their electric organs.";  
RL J. Exp. Biol. 203:665-683(2000).  
DR EMBL; AF201634; AAF43345.1; -.  
DR InterPro; IPR004321; RAG2.  
DR Pfam; PF03089; RAG2; 1.  
FT NON\_TER 1 1  
FT NON\_TER 377 377  
SQ SEQUENCE 377 AA; 41331 MW; A3755C47CD878939 CRC64;

Query Match 83.3%; Score 40; DB 13; Length 377;  
Best Local Similarity 87.5%; Pred. No. 10;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
||| ||||  
Db 103 CVLFGGRC 110

RESULT 10

Q9IAG1

ID Q9IAG1 PRELIMINARY; PRT; 377 AA.  
AC Q9IAG1;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Recombination-activating protein 2 (Fragment).  
OS *Hyperopisus bebe*.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;  
OC Osteoglossiformes; Mormyridae; *Hyperopisus*.  
OX NCBI\_TaxID=91725;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20115608; PubMed=10648209;  
RA Sullivan J.P., Lavoue S., Hopkins C.D.;  
RT "Molecular systematics of the African electric fishes (Mormyroidea:

RT Teleostei) and a model for the evolution of their electric organs.";  
 RL J. Exp. Biol. 203:665-683(2000).  
 DR EMBL; AF201633; AAF43344.1; -.  
 DR InterPro; IPR004321; RAG2.  
 DR Pfam; PF03089; RAG2; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 377 377  
 SQ SEQUENCE 377 AA; 41324 MW; C3C5A2BBE34EF6FC CRC64;

Query Match 83.3%; Score 40; DB 13; Length 377;  
 Best Local Similarity 87.5%; Pred. No. 10;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
 ||| |||  
 Db 103 CVLFGGRC 110

# RESULT 11

Q9IAE5

ID Q9IAE5 PRELIMINARY; PRT; 377 AA.  
 AC Q9IAE5;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Recombination-activating protein 2 (Fragment).  
 OS Petrocephalus microphthalmus.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;  
 OC Osteoglossiformes; Mormyridae; Petrocephalus.  
 OX NCBI\_TaxID=112157;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20115608; PubMed=10648209;  
 RA Sullivan J.P., Lavoue S., Hopkins C.D.;  
 RT "Molecular systematics of the African electric fishes (Mormyroidea:  
 RT Teleostei) and a model for the evolution of their electric organs.";  
 RL J. Exp. Biol. 203:665-683(2000).  
 DR EMBL; AF201649; AAF43360.1; -.  
 DR InterPro; IPR004321; RAG2.  
 DR Pfam; PF03089; RAG2; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 377 377  
 SQ SEQUENCE 377 AA; 41212 MW; D11CD4BDAB0099B0 CRC64;

Query Match 83.3%; Score 40; DB 13; Length 377;  
 Best Local Similarity 87.5%; Pred. No. 10;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
 ||| |||  
 Db 103 CVLFGGRC 110

# RESULT 12

Q9IAF4

ID Q9IAF4 PRELIMINARY; PRT; 377 AA.



AC Q9IAF4;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Recombination-activating protein 2 (Fragment).  
 OS *Marcusenius senegalensis*.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;  
 OC Osteoglossiformes; Mormyridae; Marcusenius.  
 OX NCBI\_TaxID=42650;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20115608; PubMed=10648209;  
 RA Sullivan J.P., Lavoue S., Hopkins C.D.;  
 RT "Molecular systematics of the African electric fishes (Mormyroidea:  
 RT Teleostei) and a model for the evolution of their electric organs.";  
 RL J. Exp. Biol. 203:665-683(2000).  
 DR EMBL; AF201640; AAF43351.1; -.  
 DR InterPro; IPR004321; RAG2.  
 DR Pfam; PF03089; RAG2; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 377 377  
 SQ SEQUENCE 377 AA; 41393 MW; A33A11B903FE33C7 CRC64;  
  
 Query Match 83.3%; Score 40; DB 13; Length 377;  
 Best Local Similarity 87.5%; Pred. No. 10;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
 ||| |||  
 Db 103 CVLFGGRC 110

# RESULT 13

Q9IAD8  
 ID Q9IAD8 PRELIMINARY; PRT; 377 AA.  
 AC Q9IAD8;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Recombination-activating protein 2 (Fragment).  
 OS *Stomatorhinus walkeri*.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;  
 OC Osteoglossiformes; Mormyridae; Stomatorhinus.  
 OX NCBI\_TaxID=112160;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20115608; PubMed=10648209;  
 RA Sullivan J.P., Lavoue S., Hopkins C.D.;  
 RT "Molecular systematics of the African electric fishes (Mormyroidea:  
 RT Teleostei) and a model for the evolution of their electric organs.";  
 RL J. Exp. Biol. 203:665-683(2000).  
 DR EMBL; AF201656; AAF43367.1; -.  
 DR InterPro; IPR004321; RAG2.  
 DR Pfam; PF03089; RAG2; 1.  
 FT NON\_TER 1 1

FT NON\_TER 377 377  
SQ SEQUENCE 377 AA; 41529 MW; 4FD1CC06990F0E2F CRC64;

Query Match 83.3%; Score 40; DB 13; Length 377;  
Best Local Similarity 87.5%; Pred. No. 10;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
||| |||  
Db 103 CVLFGGRC 110

RESULT 14

Q9IAH1

ID Q9IAH1 PRELIMINARY; PRT; 377 AA.  
AC Q9IAH1;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Recombination-activating protein 2 (Fragment).  
OS Brienomyrus sp. CU79740.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;  
OC Osteoglossiformes; Mormyridae; Brienomyrus.  
OX NCBI\_TaxID=112278;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20115608; PubMed=10648209;  
RA Sullivan J.P., Lavoue S., Hopkins C.D.;  
RT "Molecular systematics of the African electric fishes (Mormyroidea:  
RT Teleostei) and a model for the evolution of their electric organs.";  
RL J. Exp. Biol. 203:665-683(2000).  
DR EMBL; AF201621; AAF43332.1; -.  
DR InterPro; IPR004321; RAG2.  
DR Pfam; PF03089; RAG2; 1.  
FT NON\_TER 1 1  
FT NON\_TER 377 377  
SQ SEQUENCE 377 AA; 41475 MW; 735853EEA67408FE CRC64;

Query Match 83.3%; Score 40; DB 13; Length 377;  
Best Local Similarity 87.5%; Pred. No. 10;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
||| |||  
Db 103 CVLFGGRC 110

RESULT 15

Q9IAE2

ID Q9IAE2 PRELIMINARY; PRT; 377 AA.  
AC Q9IAE2;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Recombination-activating protein 2 (Fragment).  
OS Petrocephalus soudanensis.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;  
 OC Osteoglossiformes; Mormyridae; Petrocephalus.  
 OX NCBI\_TaxID=91712;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20115608; PubMed=10648209;  
 RA Sullivan J.P., Lavoue S., Hopkins C.D.;  
 RT "Molecular systematics of the African electric fishes (Mormyroidea:  
 RT Teleostei) and a model for the evolution of their electric organs.";  
 RL J. Exp. Biol. 203:665-683(2000).  
 DR EMBL; AF201652; AAF43363.1; -.  
 DR InterPro; IPR004321; RAG2.  
 DR Pfam; PF03089; RAG2; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 377 377  
 SQ SEQUENCE 377 AA; 41192 MW; 68AF50B04FFD3541 CRC64;

Query Match 83.3%; Score 40; DB 13; Length 377;  
 Best Local Similarity 87.5%; Pred. No. 10;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVLRGGRC 8  
 ||| ||||  
 Db 103 CVLFGGRC 110

Search completed: November 13, 2003, 09:51:00  
 Job time : 22.0833 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:31:40 ; Search time 30.2812 Seconds  
(without alignments)  
47.176 Million cell updates/sec

Title: US-09-228-866-5  
Perfect score: 51  
Sequence: 1 CNSRLQLRC 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_19Jun03:\*

- 1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:\*
- 2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:\*
- 3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:\*
- 4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:\*
- 5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:\*
- 6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:\*
- 7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:\*
- 8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:\*
- 9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:\*
- 10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:\*
- 11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:\*
- 12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:\*
- 13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:\*
- 14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:\*
- 15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:\*
- 16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:\*
- 17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:\*
- 18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:\*
- 19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:\*
- 20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:\*
- 21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:\*
- 22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:\*
- 23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:\*
- 24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

# SUMMARIES

Result		%					
No.	Score	Query	Match	Length	DB	ID	Description
1	51	100.0		9	18	AAW13411	Brain homing pepti
2	51	100.0		9	21	AAB07391	Brain homing pepti
3	51	100.0		9	22	AAE11797	Phage peptide #5 t
4	51	100.0		9	23	AAU10708	Brain homing pepti
5	46	90.2		9	18	AAW13410	Brain homing pepti
6	46	90.2		9	21	AAB07387	Brain homing pepti
7	46	90.2		9	22	AAE11793	Phage peptide #1 t
8	46	90.2		9	23	AAU10704	Brain homing pepti
9	46	90.2		9	24	ABU59529	Brain receptor tar
10	39	76.5		70	21	AAG36143	Arabidopsis thalia
11	39	76.5		70	21	AAG38425	Arabidopsis thalia
12	39	76.5		94	21	AAG36142	Arabidopsis thalia
13	39	76.5		94	21	AAG38424	Arabidopsis thalia
14	39	76.5		113	21	AAG38423	Arabidopsis thalia
15	38	74.5		36	22	ABG07364	Novel human diagno
16	37	72.5		468	22	ABB65114	Drosophila melanog
17	37	72.5		910	23	ABG79618	Nematode worm ribo
18	37	72.5		910	23	AAM50934	Nematode 3'-5' exo
19	36	70.6		148	22	ABB66571	Drosophila melanog
20	35	68.6		153	21	AAB43041	Human ORFX ORF2805
21	35	68.6		615	22	ABB67585	Drosophila melanog
22	34	66.7		40	23	AAM48070	RNA binding surrog
23	34	66.7		74	21	AAG45750	Arabidopsis thalia
24	34	66.7		88	23	AAO21319	Potato KCP-like pr
25	34	66.7		99	22	ABG09885	Novel human diagno
26	34	66.7		210	24	ABR41767	Human DITHP bioche
27	33	64.7		84	22	AAU29594	Novel human secret
28	33	64.7		94	22	AAO10270	Human polypeptide
29	33	64.7		97	23	AAO21325	Arabidopsis thalia
30	33	64.7		104	22	ABB68248	Drosophila melanog
31	33	64.7		142	22	AAU19527	Human diagnostic a
32	33	64.7		174	23	AAU83090	Novel secreted pro
33	33	64.7		188	16	AAR79915	Human trypsin inhi
34	33	64.7		198	16	AAR79914	Trypsin inhibitory
35	33	64.7		258	21	AAG79301	Protein encoded by
36	33	64.7		258	23	ABJ05597	Breast cancer-asso
37	33	64.7		258	23	ABG61801	Prostate cancer-as
38	33	64.7		258	23	ABG61802	Prostate cancer-as
39	33	64.7		273	23	ABP43691	Human G713 protein
40	33	64.7		341	22	ABB71537	Drosophila melanog
41	33	64.7		426	21	AAG57859	Arabidopsis thalia
42	33	64.7		430	21	AAG57858	Arabidopsis thalia
43	33	64.7		458	21	AAV90962	Human G713 protein
44	33	64.7		542	21	AAG57857	Arabidopsis thalia
45	33	64.7		2011	24	ABJ37913	NOVX protein seque

# ALIGNMENTS

RESULT 1

AAW13411

ID AAW13411 standard; Peptide; 9 AA.

XX

AC AAW13411;

XX

DT 15-JAN-1998 (first entry)

XX

DE Brain homing peptide.

XX

KW Brain homing peptide; in vivo panning; screening; phage display;  
KW drug delivery.

XX

OS Synthetic.

XX

PN WO9710507-A1.

XX

PD 20-MAR-1997.

XX

PF 10-SEP-1996; 96WO-US14600.

XX

PR 11-SEP-1995; 95US-0526710.

PR 11-SEP-1995; 95US-0526708.

XX

PA (LJOL-) LA JOLLA CANCER RES FOUND.

XX

PI Pasqualini R, Ruoslahti E;

XX

DR WPI; 1997-202359/18.

XX

PT Obtaining compound that homes to selected organ or tissue - by in  
PT vivo panning method, specifically to identify brain, kidney,  
PT angiogenic vasculature or tumour tissue homing peptide(s)

XX

PS Claim 11; Page 67; 75pp; English.

XX

CC This synthetic peptide is a claimed example of a brain-homing  
CC peptide that was identified using a novel method for obtaining  
CC molecules that home to a selected organ or tissue. This in vivo  
CC panning method typically involves administering a phage display  
CC library to a subject, and identifying expressed peptides which  
CC home to the desired organ or tissue, e.g. brain, kidney, angiogenic  
CC vascular tissue or tumour tissue. The isolated peptides (see  
CC AAW13412-52, AAW11181-86) can be used to target e.g. drugs, toxins or  
CC labels to the selected organ/tissue (claimed) or to identify and/or  
CC isolate target molecules (claimed). The peptides can be directly  
CC identified in vivo, as compared to prior art in vitro screening  
CC methods, which require further examination to see if they maintain  
CC specificity in vivo.

XX

SQ Sequence 9 AA;

Query Match 100.0%; Score 51; DB 18; Length 9;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CNSRLQLRC 9

Db                    |||||  
                      1 CNSRLQLRC 9

RESULT 2

AAB07391

ID    AAB07391 standard; peptide; 9 AA.

XX

AC    AAB07391;

XX

DT    17-OCT-2000    (first entry)

XX

DE    Brain homing peptide # 5.

XX

KW    Brain; homing peptide; organ targeting; tissue targeting; mouse; cyclic.

XX

OS    Mus sp.

XX

FH    Key                    Location/Qualifiers

FT    Disulfide-bond    1..9

FT                                /note= "Can optionally form a cyclic peptide"

XX

PN    US6068829-A.

XX

PD    30-MAY-2000.

XX

PF    23-JUN-1997;    97US-0862855.

XX

PR    11-SEP-1995;    95US-0526710.

PR    10-MAR-1997;    97US-0813273.

XX

PA    (BURN-) BURNHAM INST.

XX

PI    Pasqualini R,    Ruoslahti E;

XX

DR    WPI; 2000-410850/35.

XX

PT    Identifying and recovering organ homing molecules or peptides by in  
PT    vivo panning comprises administering a library of diverse peptides  
PT    linked to a tag which facilitates recovery of these peptides    -

XX

PS    Example 2; Column 17; 20pp; English.

XX

CC    The present sequence is a mouse brain homing peptide. This sequence was  
CC    identified by using in vivo panning to screen a library of potential  
CC    organ homing molecules. The present sequence can be used to direct a  
CC    moiety to a the brain tissue, by linking the moiety to the present  
CC    sequence. Examples of potential moieties are drugs, toxins or a  
CC    detectable label. The present sequence contains a SRL amino acid motif.

XX

SQ    Sequence    9 AA;

Query Match                    100.0%;    Score 51;    DB 21;    Length 9;

Best Local Similarity    100.0%;    Pred. No. 9.3e+05;

Matches    9;    Conservative    0;    Mismatches    0;    Indels    0;    Gaps    0;

Qy                    1 CNSRLQLRC 9

Db

|||||||  
1 CNSRLQLRC 9

RESULT 3

AAE11797

ID AAE11797 standard; peptide; 9 AA.

XX

AC AAE11797;

XX

DT 18-DEC-2001 (first entry)

XX

DE Phage peptide #5 targetted to brain.

XX

KW Enriched library fraction; brain; kidney; tumour; panning; diagnostic;  
KW molecular medicine; drug delivery; peptidomimetic; pharmaceutical.

XX

OS Bacteriophage.

XX

FH Key Location/Qualifiers

FT Domain

3..5

FT

/label= SRL\_motif

XX

PN US6296832-B1.

XX

PD 02-OCT-2001.

XX

PF 08-JAN-1999; 99US-0226985.

XX

PR 23-JUN-1997; 97US-0862855.

PR 11-SEP-1995; 95US-0526710.

PR 10-MAR-1997; 97US-0813273.

XX

PA (BURN-) BURNHAM INST.

XX

PI Ruoslahti E, Pasqualini R;

XX

DR WPI; 2001-610691/70.

XX

PT Enriched library fraction comprising molecules recovered by in vivo  
PT panning that selectively home to a selected organ or tissue useful for  
PT treating disease or in diagnostic methods -

XX

PS Example 2; Column 17; 21pp; English.

XX

CC The invention relates to an enriched library fraction containing  
CC molecules that selectively home to a selected organ or tissue such as  
CC brain, kidney or tumour recovered by in vivo panning. The invention  
CC generally relates to the field of molecular medicine, drug delivery and  
CC to a method of invivo panning for identifying a molecule that homes to a  
CC specific organ. The molecules, e.g., peptides, peptidomimetics, proteins  
CC and fragments of proteins contained in an enriched library fraction may  
CC be administered to a subject as part of a pharmaceutical composition to  
CC treat disease or in diagnostic methods. The present sequence is a  
CC peptide from bacteriophage targetted to brain.

XX

SQ Sequence 9 AA;



Query Match 100.0%; Score 51; DB 22; Length 9;  
Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CNSRLQLRC 9  
|||  
Db 1 CNSRLQLRC 9

RESULT 4

AAU10708

ID AAU10708 standard; peptide; 9 AA.

XX

AC AAU10708;

XX

DT 12-MAR-2002 (first entry)

XX

DE Brain homing peptide #5 useful for delivery of target molecules.

XX

KW Organ targeting; tissue targeting; cancer; tumour homing molecule;  
KW delivery of target molecule; brain homing peptide.

XX

OS Synthetic.

XX

PN US6306365-B1.

XX

PD 23-OCT-2001.

XX

PF 08-JAN-1999; 99US-0227906.

XX

PR 23-JUN-1997; 97US-0862855.

PR 11-SEP-1995; 95US-0526710.

PR 10-MAR-1997; 97US-0813273.

XX

PA (BURN-) BURNHAM INST.

XX

PI Ruoslahti E, Pasqualini R;

XX

DR WPI; 2002-040196/05.

XX

PT Recovering molecules that home to an organ or tissue, useful for  
PT identifying molecules that home to a specific organ or tissue, e.g.  
PT identifying a tumour homing molecule to identify the presence of cancer,  
PT by in vivo panning of a library -

XX

PS Example 2; Column 17; 21pp; English.

XX

CC The present invention relates to a method of recovering molecules that  
CC home to a selected organ or tissue. The method comprises administering  
CC to the subject the library of diverse molecules, collecting a sample of  
CC the selected organ or tissue (e.g. brain or kidney), and recovering from  
CC the sample several molecules that home to the selected organ or tissue.  
CC The method is useful for identifying molecules, particularly useful for  
CC screening large number of molecules (e.g. peptides), that home to a  
CC specific organ. The identified molecule is useful for e.g. raising an  
CC antibody specific for a target molecule, targeting a desired moiety

CC (e.g. drug, toxin or detectable label) to the selected organ.  
CC Specifically, the method is useful for identifying the presence of cancer  
CC in a subject by linking an appropriate moiety to a tumour homing  
CC molecule. The present method provides a direct means for identifying  
CC molecules that specifically home to a selected organ and, therefore  
CC provides a significant advantage over previous methods, which require  
CC that a molecule identified using an in vitro screening method  
CC subsequently be examined to determine if it maintains its specificity in  
CC vivo. AAU10704-AAU10723 represent brain homing peptides described in  
CC the present invention.

XX

SQ Sequence 9 AA;

Query Match 100.0%; Score 51; DB 23; Length 9;  
Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9

|||||

Db 1 CNSRLQLRC 9

#### RESULT 5

AAW13410

ID AAW13410 standard; Peptide; 9 AA.

XX

AC AAW13410;

XX

DT 15-JAN-1998 (first entry)

XX

DE Brain homing peptide.

XX

KW Brain homing peptide; in vivo panning; screening; phage display.

XX

OS Synthetic.

XX

PN WO9710507-A1.

XX

PD 20-MAR-1997.

XX

PF 10-SEP-1996; 96WO-US14600.

XX

PR 11-SEP-1995; 95US-0526710.

PR 11-SEP-1995; 95US-0526708.

XX

PA (LJOL-) LA JOLLA CANCER RES FOUND.

XX

PI Pasqualini R, Ruoslahti E;

XX

DR WPI; 1997-202359/18.

XX

PT Obtaining compound that homes to selected organ or tissue - by in  
PT vivo panning method, specifically to identify brain, kidney,  
PT angiogenic vasculature or tumour tissue homing peptide(s)

XX

PS Claim 11; Page 67; 75pp; English.

XX

CC This synthetic peptide is a claimed example of a brain-homing  
 CC peptide that was identified using a novel method for obtaining  
 CC molecules that home to a selected organ or tissue. This in vivo  
 CC panning method typically involves administering a phage display  
 CC library to a subject, and identifying expressed peptides which  
 CC home to the desired organ or tissue, e.g. brain, kidney, angiogenic  
 CC vascular tissue or tumour tissue. The isolated peptides (see  
 CC AAW13412-52, AAW11181-86) can be used to target e.g. drugs, toxins or  
 CC labels to the selected organ/tissue (claimed) or to identify and/or  
 CC isolate target molecules (claimed). The peptides can be directly  
 CC identified in vivo, as compared to prior art in vitro screening  
 CC methods, which require further examination to see if they maintain  
 CC specificity in vivo.

XX

SQ Sequence 9 AA;

Query Match 90.2%; Score 46; DB 18; Length 9;  
 Best Local Similarity 88.9%; Pred. No. 9.3e+05;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CNSRLQLRC 9  
 ||||| |||  
 Db 1 CNSRLHLRC 9

# RESULT 6

AAB07387

ID AAB07387 standard; peptide; 9 AA.

XX

AC AAB07387;

XX

DT 17-OCT-2000 (first entry)

XX

DE Brain homing peptide # 1.

XX

KW Brain; homing peptide; organ targeting; tissue targeting; mouse; cyclic.

XX

OS Mus sp.

XX

FH Key Location/Qualifiers

FT Disulfide-bond 1..9

FT /note= "Can optionally form a cyclic peptide"

XX

PN US6068829-A.

XX

PD 30-MAY-2000.

XX

PF 23-JUN-1997; 97US-0862855.

XX

PR 11-SEP-1995; 95US-0526710.

PR 10-MAR-1997; 97US-0813273.

XX

PA (BURN-) BURNHAM INST.

XX

PI Pasqualini R, Ruoslahti E;

XX

DR WPI; 2000-410850/35.

XX  
PT Identifying and recovering organ homing molecules or peptides by in  
PT vivo panning comprises administering a library of diverse peptides  
PT linked to a tag which facilitates recovery of these peptides -  
XX  
PS Example 2; Column 17; 20pp; English.  
XX  
CC The present sequence is a mouse brain homing peptide. This sequence was  
CC identified by using in vivo panning to screen a library of potential  
CC organ homing molecules. The present sequence can be used to direct a  
CC moiety to a the brain tissue, by linking the moiety to the present  
CC sequence. Examples of potential moieties are drugs, toxins or a  
CC detectable label. The present sequence contains a SRL amino acid motif.  
XX  
SQ Sequence 9 AA;

Query Match 90.2%; Score 46; DB 21; Length 9;  
Best Local Similarity 88.9%; Pred. No. 9.3e+05;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CNSRLQLRC 9  
||| |||  
Db 1 CNSRLHLRC 9

RESULT 7  
AAE11793  
ID AAE11793 standard; peptide; 9 AA.  
XX  
AC AAE11793;  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DE Phage peptide #1 targetted to brain.  
XX  
KW Enriched library fraction; brain; kidney; tumour; panning; diagnostic;  
KW molecular medicine; drug delivery; peptidomimetic; pharmaceutical.  
XX  
OS Bacteriophage.  
XX  
FH Key Location/Qualifiers  
FT Domain 3..5  
FT /label= SRL\_motif  
XX  
PN US6296832-B1.  
XX  
PD 02-OCT-2001.  
XX  
PF 08-JAN-1999; 99US-0226985.  
XX  
PR 23-JUN-1997; 97US-0862855.  
PR 11-SEP-1995; 95US-0526710.  
PR 10-MAR-1997; 97US-0813273.  
XX  
PA (BURN-) BURNHAM INST.  
XX  
PI Ruoslahti E, Pasqualini R;

XX  
 DR WPI; 2001-610691/70.  
 XX  
 PT Enriched library fraction comprising molecules recovered by in vivo  
 PT panning that selectively home to a selected organ or tissue useful for  
 PT treating disease or in diagnostic methods -  
 XX  
 PS Example 2; Column 17; 21pp; English.  
 XX  
 CC The invention relates to an enriched library fraction containing  
 CC molecules that selectively home to a selected organ or tissue such as  
 CC brain, kidney or tumour recovered by in vivo panning. The invention  
 CC generally relates to the field of molecular medicine, drug delivery and  
 CC to a method of invivo panning for identifying a molecule that homes to a  
 CC specific organ. The molecules, e.g., peptides, peptidomimetics, proteins  
 CC and fragments of proteins contained in an enriched library fraction may  
 CC be administered to a subject as part of a pharmaceutical composition to  
 CC treat disease or in diagnostic methods. The present sequence is a  
 CC peptide from bacteriophage targetted to brain.  
 XX  
 SQ Sequence 9 AA;

Query Match 90.2%; Score 46; DB 22; Length 9;  
 Best Local Similarity 88.9%; Pred. No. 9.3e+05;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
 ||||| |||  
 Db 1 CNSRLHLRC 9

# RESULT 8

AAU10704

ID AAU10704 standard; peptide; 9 AA.

XX

AC AAU10704;

XX

DT 12-MAR-2002 (first entry)

XX

DE Brain homing peptide #1 useful for delivery of target molecules.

XX

KW Organ targeting; tissue targeting; cancer; tumour homing molecule;  
 KW delivery of target molecule; brain homing peptide.

XX

OS Synthetic.

XX

PN US6306365-B1.

XX

PD 23-OCT-2001.

XX

PF 08-JAN-1999; 99US-0227906.

XX

PR 23-JUN-1997; 97US-0862855.

PR 11-SEP-1995; 95US-0526710.

PR 10-MAR-1997; 97US-0813273.

XX

PA (BURN-) BURNHAM INST.

XX  
 PI Ruoslahti E, Pasqualini R;  
 XX  
 DR WPI; 2002-040196/05.  
 XX  
 PT Recovering molecules that home to an organ or tissue, useful for  
 PT identifying molecules that home to a specific organ or tissue, e.g.  
 PT identifying a tumour homing molecule to identify the presence of cancer,  
 PT by in vivo panning of a library -  
 XX  
 PS Example 2; Column 17; 21pp; English.  
 XX  
 CC The present invention relates to a method of recovering molecules that  
 CC home to a selected organ or tissue. The method comprises administering  
 CC to the subject the library of diverse molecules, collecting a sample of  
 CC the selected organ or tissue (e.g. brain or kidney), and recovering from  
 CC the sample several molecules that home to the selected organ or tissue.  
 CC The method is useful for identifying molecules, particularly useful for  
 CC screening large number of molecules (e.g. peptides), that home to a  
 CC specific organ. The identified molecule is useful for e.g. raising an  
 CC antibody specific for a target molecule, targeting a desired moiety  
 CC (e.g. drug, toxin or detectable label) to the selected organ.  
 CC Specifically, the method is useful for identifying the presence of cancer  
 CC in a subject by linking an appropriate moiety to a tumour homing  
 CC molecule. The present method provides a direct means for identifying  
 CC molecules that specifically home to a selected organ and, therefore  
 CC provides a significant advantage over previous methods, which require  
 CC that a molecule identified using an in vitro screening method  
 CC subsequently be examined to determine if it maintains its specificity in  
 CC vivo. AAU10704-AAU10723 represent brain homing peptides described in  
 CC the present invention.  
 XX  
 SQ Sequence 9 AA;

Query Match 90.2%; Score 46; DB 23; Length 9;  
 Best Local Similarity 88.9%; Pred. No. 9.3e+05;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CNSRLQLRC 9  
 ||||| |||  
 Db 1 CNSRLHLRC 9

# RESULT 9

ABU59529

ID ABU59529 standard; Peptide; 9 AA.

XX

AC ABU59529;

XX

DT 22-APR-2003 (first entry)

XX

DE Brain receptor targeting peptide #1.

XX

KW Targeting ligand; bioactive agent; polymer matrix; cancer; cytostatic;  
 KW cathepsin-D substrate; peptides; neuroreceptor; adrenal receptor;  
 KW fibronectin; vitronectin; integrin; RGD motif; angiogenic endothelium;  
 KW tumour; cationic cancer-targeting peptide.

XX  
 OS Synthetic.  
 XX  
 PN US2002041898-A1.  
 XX  
 PD 11-APR-2002.  
 XX  
 PF 25-JUL-2001; 2001US-0912609.  
 XX  
 PR 05-JAN-2000; 2000US-0478124.  
 PR 31-OCT-2000; 2000US-0703474.  
 XX  
 PA (UNGE/) UNGER E C.  
 PA (MATS/) MATSUNAGA T O.  
 PA (RAMA/) RAMASWAMI V.  
 PA (ROMA/) ROMANOWSKI M J.  
 XX  
 PI Unger EC, Matsunaga TO, Ramaswami V, Romanowski MJ;  
 XX  
 DR WPI; 2003-208921/20.  
 XX  
 PT Targeted delivery system comprising a bioactive agent homogeneously  
 PT dispersed in a targeted matrix is especially useful in cancer therapy  
 PT -  
 XX  
 PS Claim 23; Page 37; 46pp; English.  
 XX  
 CC The invention relates to a composition comprising a bioactive agent  
 CC homogeneously dispersed in a targeted matrix (polymer and targeting  
 CC ligand). Also included are a targeted matrix for use as a delivery  
 CC vehicle comprising a polymer associated with a targeting ligand,  
 CC enhancing the bioavailability of an agent comprising administration  
 CC of the composition and treating cancer comprising administration of the  
 CC novel composition. The method is useful for targeted delivery of a drug,  
 CC especially in cancer therapy. The targeting ligand may be a peptide.  
 CC Examples of targeting peptides are disclosed including cathepsin-D  
 CC substrate peptides, peptides targeting receptors in the brain and  
 CC kidney, peptides recognising fibronectin- and vitronectin-binding  
 CC integrins, peptides targeting the RGD (Arg-Gly-Asp)-motif in, e.g.,  
 CC antibodies, peptides targeting the angiogenic endothelium of solid  
 CC tumours, tissue specific peptides (e.g. of lung, skin, pancreas,  
 CC intestine, uterus, adrenal gland and retina), and cationic cancer-  
 CC targeting peptides. The present sequence is a peptide targeting  
 CC ligand disclosed in the invention.  
 XX  
 SQ Sequence 9 AA;  
  
 Query Match 90.2%; Score 46; DB 24; Length 9;  
 Best Local Similarity 88.9%; Pred. No. 9.3e+05;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
 Qy 1 CNSRLQLRC 9  
 ||||| |||  
 Db 1 CNSRLHLRC 9

RESULT 10

AAG36143

ID AAG36143 standard; Protein; 70 AA.

XX

AC AAG36143;

XX

DT 18-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 44252.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX

OS Arabidopsis thaliana.

XX

PN EP1033405-A2.

XX

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-0301439.

XX

PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.



PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.

PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.

PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 76.5%; Score 39; DB 21; Length 70;  
 Best Local Similarity 77.8%; Pred. No. 12;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
 |||| | ||  
 Db 12 CNSRCQERC 20

# RESULT 11

AAG38425

ID AAG38425 standard; Protein; 70 AA.

XX

AC AAG38425;

XX

DT 18-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 47404.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.

XX

OS Arabidopsis thaliana.

XX

PN EP1033405-A2.

XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.

PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.

PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.

PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 76.5%; Score 39; DB 21; Length 70;  
 Best Local Similarity 77.8%; Pred. No. 12;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
 |||| | ||  
 Db 12 CNSRCQERC 20

# RESULT 12

AAG36142

ID AAG36142 standard; Protein; 94 AA.

XX

AC AAG36142;

XX

DT 18-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 44251.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.

XX

OS Arabidopsis thaliana.

XX

PN EP1033405-A2.

XX

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-0301439.

XX

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	23-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	06-MAY-1999;	99US-0132487.
PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.



PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.

PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match

76.5%; Score 39; DB 21; Length 94;

Best Local Similarity 77.8%; Pred. No. 16;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
||| |  
Db 36 CNSRCQERC 44

RESULT 13

AAG38424

ID AAG38424 standard; Protein; 94 AA.

XX

AC AAG38424;

XX

DT 18-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 47403.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX

OS Arabidopsis thaliana.

XX

PN EP1033405-A2.

XX

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-0301439.

XX

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.

PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.

PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 76.5%; Score 39; DB 21; Length 94;  
 Best Local Similarity 77.8%; Pred. No. 16;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
 |||| | ||  
 Db 36 CNSRCQERC 44

RESULT 14  
 AAG38423  
 ID AAG38423 standard; Protein; 113 AA.  
 XX  
 AC AAG38423;  
 XX  
 DT 18-OCT-2000 (first entry)  
 XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 47402.  
 XX  
 KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 PN EP1033405-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 25-FEB-2000; 2000EP-0301439.  
 XX  
 PR 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 23-APR-1999; 99US-0130891.  
 PR 28-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 06-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.

PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.



PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.

PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 76.5%; Score 39; DB 21; Length 113;  
 Best Local Similarity 77.8%; Pred. No. 18;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
 |||||  
 Db 55 CNSRCQERC 63

# RESULT 15

ABG07364

ID ABG07364 standard; Protein; 36 AA.

XX

AC ABG07364;

XX

DT 13-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #7355.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US08631.

XX

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX

PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS71551.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 20; SEQ ID No 37723; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 36 AA;

Query Match 74.5%; Score 38; DB 22; Length 36;  
 Best Local Similarity 77.8%; Pred. No. 9.8;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
 | ||| |||  
 Db 21 CQSRLLLRC 29

Search completed: November 13, 2003, 09:45:25  
 Job time : 31.2812 secs

GenCore version 5.1.6  
 Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:45:35 ; Search time 18.6562 Seconds

(without alignments)  
88.069 Million cell updates/sec

Title: US-09-228-866-5  
Perfect score: 51  
Sequence: 1 CNSRLQLRC 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	%		DB	ID	Description
		Query	Match Length			
1	46	90.2	9	12	US-10-306-878-11	Sequence 11, Appl
2	37	72.5	910	12	US-09-896-186B-16	Sequence 16, Appl
3	35	68.6	1154	12	US-10-100-818-4	Sequence 4, Appli
4	34	66.7	40	11	US-09-852-455-53	Sequence 53, Appl
5	34	66.7	88	10	US-09-950-933A-74	Sequence 74, Appl
6	33	64.7	97	10	US-09-950-933A-80	Sequence 80, Appl
7	33	64.7	117	11	US-09-789-390-41	Sequence 41, Appl

8	33	64.7	174	10	US-09-893-737-32	Sequence 32, Appl
9	33	64.7	191	11	US-09-789-390-42	Sequence 42, Appl
10	33	64.7	191	11	US-09-789-390-46	Sequence 46, Appl
11	33	64.7	239	15	US-10-091-135-85	Sequence 85, Appl
12	33	64.7	257	11	US-09-800-198-88	Sequence 88, Appl
13	33	64.7	258	10	US-09-808-602-110	Sequence 110, App
14	33	64.7	258	11	US-09-800-198-96	Sequence 96, Appl
15	33	64.7	458	9	US-09-416-384A-5	Sequence 5, Appli
16	32.5	63.7	28	10	US-09-934-060A-20	Sequence 20, Appl
17	32.5	63.7	556	10	US-09-934-060A-6	Sequence 6, Appli
18	32	62.7	371	11	US-09-975-719-295	Sequence 295, App
19	31	60.8	9	9	US-09-760-599-20	Sequence 20, Appl
20	31	60.8	96	11	US-09-764-891-5362	Sequence 5362, Ap
21	31	60.8	172	15	US-10-278-173-84	Sequence 84, Appl
22	31	60.8	196	10	US-09-925-300-1075	Sequence 1075, Ap
23	31	60.8	403	12	US-10-287-274-329	Sequence 329, App
24	31	60.8	437	15	US-10-156-761-11265	Sequence 11265, A
25	31	60.8	455	15	US-10-156-761-11516	Sequence 11516, A
26	31	60.8	569	11	US-09-805-337A-2	Sequence 2, Appli
27	31	60.8	760	11	US-09-759-130B-440	Sequence 440, App
28	31	60.8	760	11	US-09-759-130B-446	Sequence 446, App
29	31	60.8	760	12	US-10-190-115-36	Sequence 36, Appl
30	31	60.8	760	14	US-10-042-431-70	Sequence 70, Appl
31	31	60.8	760	14	US-10-042-431-76	Sequence 76, Appl
32	31	60.8	1400	12	US-10-354-358-42	Sequence 42, Appl
33	31	60.8	1400	15	US-10-123-036-4	Sequence 4, Appli
34	30	58.8	9	9	US-09-760-599-4	Sequence 4, Appli
35	30	58.8	9	9	US-09-760-599-28	Sequence 28, Appl
36	30	58.8	38	9	US-09-925-299-1360	Sequence 1360, Ap
37	30	58.8	38	11	US-09-925-299-1360	Sequence 1360, Ap
38	30	58.8	51	14	US-10-011-445-65	Sequence 65, Appl
39	30	58.8	75	15	US-10-128-714-8326	Sequence 8326, Ap
40	30	58.8	95	15	US-10-128-714-3326	Sequence 3326, Ap
41	30	58.8	100	10	US-09-950-933A-40	Sequence 40, Appl
42	30	58.8	115	10	US-09-950-933A-65	Sequence 65, Appl
43	30	58.8	222	12	US-10-259-165-44	Sequence 44, Appl
44	30	58.8	222	12	US-10-259-165-392	Sequence 392, App
45	30	58.8	245	15	US-10-125-001-14	Sequence 14, Appl

#### ALIGNMENTS

#### RESULT 1

US-10-306-878-11

; Sequence 11, Application US/10306878

; Publication No. US20030175819A1

; GENERAL INFORMATION:

; APPLICANT: Reed, John C.

; APPLICANT: Guo, Bin

; TITLE OF INVENTION: Methods for Identifying Modulators of

; TITLE OF INVENTION: Apoptosis

; FILE REFERENCE: P-LJ 5535

; CURRENT APPLICATION NUMBER: US/10/306,878

; CURRENT FILING DATE: 2002-11-27

; PRIOR APPLICATION NUMBER: US 60/334,149

; PRIOR FILING DATE: 2001-11-28

; NUMBER OF SEQ ID NOS: 28  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 11  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic construct  
US-10-306-878-11

Query Match 90.2%; Score 46; DB 12; Length 9;  
Best Local Similarity 88.9%; Pred. No. 6e+05;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
      |||||  
Db 1 CNSRLHLRC 9

#### RESULT 2

US-09-896-186B-16  
; Sequence 16, Application US/09896186B  
; Publication No. US20030166227A1  
; GENERAL INFORMATION:  
; APPLICANT: Joshua Z. Levin  
; APPLICANT: Ken Phillips  
; APPLICANT: Greg Budziszewski  
; APPLICANT: Fred Meins  
; APPLICANT: Zhenya Glazov  
; TITLE OF INVENTION: Methods of Controlling Gene Expression  
; FILE REFERENCE: PB/5-31481A  
; CURRENT APPLICATION NUMBER: US/09/896,186B  
; CURRENT FILING DATE: 2002-04-04  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 16  
; LENGTH: 910  
; TYPE: PRT  
; ORGANISM: C. elegans  
US-09-896-186B-16

Query Match 72.5%; Score 37; DB 12; Length 910;  
Best Local Similarity 75.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CNSRLQLR 8  
      |||||:  
Db 771 CNSRLQIK 778

#### RESULT 3

US-10-100-818-4  
; Sequence 4, Application US/10100818  
; Publication No. US20030176333A1  
; GENERAL INFORMATION:  
; APPLICANT: Lorens, James B.  
; APPLICANT: Xu, Weiduan

```
; APPLICANT: Bogenberger, Jakob
; APPLICANT: Rigel Pharmaceuticals, Incorporated
; TITLE OF INVENTION: CASPR3: Modulators of Angiogenesis
; FILE REFERENCE: 021044-001900US
; CURRENT APPLICATION NUMBER: US/10/100,818
; CURRENT FILING DATE: 2002-03-18
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1154
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: full length contactin associated protein 3
; OTHER INFORMATION: (CASPR3)
US-10-100-818-4
```

```
Query Match          68.6%; Score 35; DB 12; Length 1154;
Best Local Similarity 66.7%; Pred. No. 4.9e+02;
Matches      6; Conservative    0; Mismatches    3; Indels      0; Gaps      0;
```

```
Qy      1 CNSRLQLRC 9
          |  |  |  |
Db      542 CEQRLALRC 550
```

#### RESULT 4

```
US-09-852-455-53
; Sequence 53, Application US/09852455
; Publication No. US20030054348A1
; GENERAL INFORMATION:
; APPLICANT: BLUME, ARTHUR J.
; APPLICANT: GOLDSTEIN, NEIL
; APPLICANT: PILLUTA, RENUKA
; APPLICANT: HSIAO, KU-CHUAN
; APPLICANT: PRENDERGAST, JOHN
; TITLE OF INVENTION: METHODS OF IDENTIFYING THE ACTIVITY OF GENE PRODUCTS
; FILE REFERENCE: 2598-4004US1
; CURRENT APPLICATION NUMBER: US/09/852,455
; CURRENT FILING DATE: 2001-05-09
; PRIOR APPLICATION NUMBER: 60/202,912
; PRIOR FILING DATE: 2000-05-09
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 53
; LENGTH: 40
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: peptide
US-09-852-455-53
```

```
Query Match          66.7%; Score 34; DB 11; Length 40;
Best Local Similarity 55.6%; Pred. No. 32;
Matches      5; Conservative    2; Mismatches    2; Indels      0; Gaps      0;
```

Qy 1 CNSRLQLRC 9  
| | | : : | |  
Db 9 CTSRVRFRC 17

RESULT 5

US-09-950-933A-74

; Sequence 74, Application US/09950933A  
; Patent No. US20020166141A1  
; GENERAL INFORMATION:  
; APPLICANT: Simmons, Carl R.  
; APPLICANT: Navarro, Pedro  
; TITLE OF INVENTION: Antimicrobial Peptides and Methods of  
; TITLE OF INVENTION: Use  
; FILE REFERENCE: 35718/238472  
; CURRENT APPLICATION NUMBER: US/09/950,933A  
; CURRENT FILING DATE: 2001-09-11  
; PRIOR APPLICATION NUMBER: 60/232,569  
; PRIOR FILING DATE: 2000-09-13  
; NUMBER OF SEQ ID NOS: 99  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 74  
; LENGTH: 88  
; TYPE: PRT  
; ORGANISM: Solanum tuberosum  
US-09-950-933A-74

Query Match 66.7%; Score 34; DB 10; Length 88;  
Best Local Similarity 55.6%; Pred. No. 67;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
| : | : : | | |  
Db 30 CDSKCKLRC 38

RESULT 6

US-09-950-933A-80

; Sequence 80, Application US/09950933A  
; Patent No. US20020166141A1  
; GENERAL INFORMATION:  
; APPLICANT: Simmons, Carl R.  
; APPLICANT: Navarro, Pedro  
; TITLE OF INVENTION: Antimicrobial Peptides and Methods of  
; TITLE OF INVENTION: Use  
; FILE REFERENCE: 35718/238472  
; CURRENT APPLICATION NUMBER: US/09/950,933A  
; CURRENT FILING DATE: 2001-09-11  
; PRIOR APPLICATION NUMBER: 60/232,569  
; PRIOR FILING DATE: 2000-09-13  
; NUMBER OF SEQ ID NOS: 99  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 80  
; LENGTH: 97  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
US-09-950-933A-80



Query Match 64.7%; Score 33; DB 10; Length 97;  
Best Local Similarity 55.6%; Pred. No. 1.1e+02;  
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
|||: ||  
Db 39 CNSKCSYRC 47

RESULT 7

US-09-789-390-41  
; Sequence 41, Application US/09789390  
; Publication No. US20030059768A1  
; GENERAL INFORMATION:  
; APPLICANT: Vernet, Corine  
; APPLICANT: Fernandes, Elma  
; APPLICANT: MacDougall, John  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Spaderna, Steven K  
; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME  
; FILE REFERENCE: 15966-692  
; CURRENT APPLICATION NUMBER: US/09/789,390  
; CURRENT FILING DATE: 2001-02-23  
; PRIOR APPLICATION NUMBER: 60/185,548  
; PRIOR FILING DATE: 2000-02-28  
; PRIOR APPLICATION NUMBER: 60/199,957  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: 60/184,951  
; PRIOR FILING DATE: 2000-02-25  
; PRIOR APPLICATION NUMBER: 60/185,967  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: 60/197,723  
; PRIOR FILING DATE: 2000-04-18  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 41  
; LENGTH: 117  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-789-390-41

Query Match 64.7%; Score 33; DB 11; Length 117;  
Best Local Similarity 55.6%; Pred. No. 1.3e+02;  
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
|| | : ||  
Db 28 CNPRCPMRC 36

RESULT 8

US-09-893-737-32  
; Sequence 32, Application US/09893737  
; Patent No. US20020110855A1  
; GENERAL INFORMATION:  
; APPLICANT: Sheppard, Paul O.

```
; APPLICANT: Presnell, Scott R.
; TITLE OF INVENTION: MAMMALIAN SECRETED PROTEINS
; FILE REFERENCE: 00-41
; CURRENT APPLICATION NUMBER: US/09/893,737
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/215,446
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 329
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 174
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-893-737-32
```

```
Query Match          64.7%;  Score 33;  DB 10;  Length 174;
Best Local Similarity 62.5%;  Pred. No. 1.9e+02;
Matches      5;  Conservative      3;  Mismatches      0;  Indels      0;  Gaps      0;
```

```
Qy      1 CNSRLQLR 8
        |||:|::|
Db      150 CNSKLRIR 157
```

#### RESULT 9

```
US-09-789-390-42
; Sequence 42, Application US/09789390
; Publication No. US20030059768A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine
; APPLICANT: Fernandes, Elma
; APPLICANT: MacDougall, John
; APPLICANT: Shimkets, Richard A
; APPLICANT: Spaderna, Steven K
; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 15966-692
; CURRENT APPLICATION NUMBER: US/09/789,390
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185,548
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: 60/199,957
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/184,951
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: 60/185,967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/197,723
; PRIOR FILING DATE: 2000-04-18
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 42
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-789-390-42
```

```
Query Match          64.7%;  Score 33;  DB 11;  Length 191;
```

Best Local Similarity 55.6%; Pred. No. 2.1e+02;  
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
||| :||  
Db 102 CNPRCPMRC 110

RESULT 10

US-09-789-390-46  
; Sequence 46, Application US/09789390  
; Publication No. US20030059768A1  
; GENERAL INFORMATION:  
; APPLICANT: Vernet, Corine  
; APPLICANT: Fernandes, Elma  
; APPLICANT: MacDougall, John  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Spaderna, Steven K  
; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME  
; FILE REFERENCE: 15966-692  
; CURRENT APPLICATION NUMBER: US/09/789,390  
; CURRENT FILING DATE: 2001-02-23  
; PRIOR APPLICATION NUMBER: 60/185,548  
; PRIOR FILING DATE: 2000-02-28  
; PRIOR APPLICATION NUMBER: 60/199,957  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: 60/184,951  
; PRIOR FILING DATE: 2000-02-25  
; PRIOR APPLICATION NUMBER: 60/185,967  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: 60/197,723  
; PRIOR FILING DATE: 2000-04-18  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 46  
; LENGTH: 191  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-789-390-46

Query Match 64.7%; Score 33; DB 11; Length 191;  
Best Local Similarity 55.6%; Pred. No. 2.1e+02;  
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
||| :||  
Db 102 CNPRCPMRC 110

RESULT 11

US-10-091-135-85  
; Sequence 85, Application US/10091135  
; Publication No. US20030039660A1  
; GENERAL INFORMATION:  
; APPLICANT: King, Te Piao  
; APPLICANT: Spangfort, Michael Dho  
; TITLE OF INVENTION: RECOMBINANT HYBRID ALLERGEN CONSTRUCTS WITH REDUCED

; TITLE OF INVENTION: ALLERGENICITY THAT RETAIN IMMUNOGENICITY OF THE NATURAL ALLERGEN

; FILE REFERENCE: 2313/1H587-US1  
; CURRENT APPLICATION NUMBER: US/10/091,135  
; CURRENT FILING DATE: 2002-03-04  
; PRIOR APPLICATION NUMBER: US 60/272,818  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 98  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 85  
; LENGTH: 239  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-091-135-85

Query Match 64.7%; Score 33; DB 15; Length 239;  
Best Local Similarity 55.6%; Pred. No. 2.6e+02;  
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
||| :||  
Db 138 CNPRCPMRC 146

#### RESULT 12

US-09-800-198-88

; Sequence 88, Application US/09800198  
; Publication No. US20030087816A1  
; GENERAL INFORMATION:  
; APPLICANT: Vernet, Cornie AM  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Herrmann, John L  
; APPLICANT: Majumder, Kumud  
; APPLICANT: Mishra, Vishna  
; APPLICANT: Mezes, Peter S  
; APPLICANT: Rastelli, Luca  
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME  
; FILE REFERENCE: 15966-697  
; CURRENT APPLICATION NUMBER: US/09/800,198  
; CURRENT FILING DATE: 2001-03-05  
; PRIOR APPLICATION NUMBER: 60/186,596  
; PRIOR FILING DATE: 2000-03-03  
; NUMBER OF SEQ ID NOS: 98  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 88  
; LENGTH: 257  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-800-198-88

Query Match 64.7%; Score 33; DB 11; Length 257;  
Best Local Similarity 55.6%; Pred. No. 2.7e+02;  
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
||| :||

Db 157 CNPRCPMRC 165

RESULT 13

US-09-808-602-110

; Sequence 110, Application US/09808602  
; Patent No. US20020155115A1  
; GENERAL INFORMATION:  
; APPLICANT: Vernet, Corine A  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Herrman, John L  
; APPLICANT: Majumder, Kumud  
; APPLICANT: Mishra, Vishnu  
; APPLICANT: Mezes, Peter S  
; APPLICANT: MacDougall, John  
; TITLE OF INVENTION: No. US20020155115A1el Proteins and Nuclec Acids Encoding  
Same  
; FILE REFERENCE: 15966-697 CIP  
; CURRENT APPLICATION NUMBER: US/09/808,602  
; CURRENT FILING DATE: 2001-03-14  
; PRIOR APPLICATION NUMBER: 09/800,198  
; PRIOR FILING DATE: 2001-03-05  
; PRIOR APPLICATION NUMBER: 60/186,596  
; PRIOR FILING DATE: 2000-03-03  
; NUMBER OF SEQ ID NOS: 114  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 110  
; LENGTH: 258  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-808-602-110

Query Match 64.7%; Score 33; DB 10; Length 258;  
Best Local Similarity 55.6%; Pred. No. 2.7e+02;  
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
||| :||  
Db 157 CNPRCPMRC 165

RESULT 14

US-09-800-198-96

; Sequence 96, Application US/09800198  
; Publication No. US20030087816A1  
; GENERAL INFORMATION:  
; APPLICANT: Vernet, Cornie AM  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Herrmann, John L  
; APPLICANT: Majumder, Kumud  
; APPLICANT: Mishra, Vishna  
; APPLICANT: Mezes, Peter S  
; APPLICANT: Rastelli, Luca  
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME  
; FILE REFERENCE: 15966-697

; CURRENT APPLICATION NUMBER: US/09/800,198  
; CURRENT FILING DATE: 2001-03-05  
; PRIOR APPLICATION NUMBER: 60/186,596  
; PRIOR FILING DATE: 2000-03-03  
; NUMBER OF SEQ ID NOS: 98  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 96  
; LENGTH: 258  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-800-198-96

Query Match 64.7%; Score 33; DB 11; Length 258;  
Best Local Similarity 55.6%; Pred. No. 2.7e+02;  
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
||| :||  
Db 157 CNPRCPMRC 165

#### RESULT 15

US-09-416-384A-5  
; Sequence 5, Application US/09416384A  
; Patent No. US20020081584A1  
; GENERAL INFORMATION:  
; APPLICANT: BLUMENFELD, Marta  
; APPLICANT: BOUGUELERET, Lydie  
; APPLICANT: CHUMAKOV, Ilya  
; APPLICANT: COHEN, Daniel  
; APPLICANT: ESSIUX, Laurent  
; TITLE OF INVENTION: Genes, proteins and biallelic markers related to  
central...  
; FILE REFERENCE: GENSET.045AUS  
; CURRENT FILING DATE: 1999-10-12  
; CURRENT APPLICATION NUMBER: US/09/416,384A  
; PRIOR APPLICATION NUMBER: 60/106,457  
; PRIOR FILING DATE: 1999-10-30  
; PRIOR APPLICATION NUMBER: 60/103,955  
; PRIOR FILING DATE: 1998-10-12  
; PRIOR APPLICATION NUMBER: 60/132,277  
; PRIOR FILING DATE: 1999-05-03  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: Patent.pm  
; SEQ ID NO 5  
; LENGTH: 458  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-416-384A-5

Query Match 64.7%; Score 33; DB 9; Length 458;  
Best Local Similarity 85.7%; Pred. No. 4.7e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CNSRLQL 7  
||||| :|  
Db 413 CNSRLKL 419

Search completed: November 13, 2003, 09:58:27  
Job time : 18.6562 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:38:30 ; Search time 9.375 Seconds  
(without alignments)  
92.322 Million cell updates/sec

Title: US-09-228-866-5  
Perfect score: 51  
Sequence: 1 CNSRLQLRC 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_76:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result	Score	% Query	Match	Length	DB	ID	Description
No.							
1	37	72.5	910	2	S40930		hypothetical prote
2	36	70.6	155	2	S50367		hypothetical prote
3	35	68.6	75	2	T44224		hypothetical prote
4	35	68.6	340	2	T22010		hypothetical prote
5	34	66.7	187	2	S62511		probable peptide m
6	34	66.7	227	2	S34220		hypothetical prote
7	34	66.7	318	2	S27977		cuticle collagen d
8	34	66.7	653	2	G96675		hypothetical prote
9	34	66.7	684	2	C96596		hypothetical prote

10	34	66.7	850	2	T14450	serine/threonine k
11	34	66.7	989	2	T01519	hypothetical prote
12	34	66.7	2652	1	VFIHB2	genome polyprotein
13	33	64.7	97	2	S71371	gibberellin-regula
14	33	64.7	736	2	T19366	hypothetical prote
15	33	64.7	1101	2	T16840	hypothetical prote
16	33	64.7	2109	2	I38414	transcription fact
17	32	62.7	173	2	S23411	FUN53 protein - ye
18	32	62.7	251	2	H83447	hypothetical prote
19	32	62.7	276	2	T24440	hypothetical prote
20	32	62.7	386	2	A35693	NADH2 dehydrogenas
21	32	62.7	386	2	F82721	two-component syst
22	32	62.7	427	2	A55353	11beta-hydroxyster
23	32	62.7	482	2	A30198	dihydrolipoamide S
24	32	62.7	525	2	T10574	hypothetical prote
25	32	62.7	648	2	B81686	conserved hypothet
26	32	62.7	651	2	E85024	probable CHP-rich
27	32	62.7	652	2	G85024	probable CHP-rich
28	32	62.7	659	2	T01520	hypothetical prote
29	32	62.7	667	2	T01999	hypothetical prote
30	32	62.7	765	2	T40674	protein transport
31	32	62.7	2731	1	VFIHJH	genome polyprotein
32	32	62.7	2733	2	S15760	genome polyprotein
33	31.5	61.8	479	1	VBEBF2	glycoprotein F - h
34	31	60.8	73	2	AG1852	hypothetical prote
35	31	60.8	104	2	T24495	hypothetical prote
36	31	60.8	111	2	G72615	hypothetical prote
37	31	60.8	129	2	G83695	hypothetical prote
38	31	60.8	133	2	AD2227	transposase all337
39	31	60.8	133	2	AF2488	transposase alr708
40	31	60.8	159	2	F86429	protein F26G16.9 [
41	31	60.8	172	2	T50694	transcription fact
42	31	60.8	199	2	T47716	transcription fact
43	31	60.8	200	2	G84822	probable LIM-domai
44	31	60.8	225	2	T33114	hypothetical prote
45	31	60.8	228	2	C90095	hypothetical prote

#### ALIGNMENTS

##### RESULT 1

S40930

hypothetical protein ZK1098.8 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 09-Sep-1997

C;Accession: S40930

R;Thomas, K.

submitted to the EMBL Data Library, February 1992

A;Reference number: S40923

A;Accession: S40930

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-910 <THO>

A;Cross-references: EMBL:Z22176; NID:g297978; PID:g297986

C;Genetics:

A;Introns: 64/1; 336/2; 382/2; 447/2; 681/2; 810/1; 852/2



Query Match 72.5%; Score 37; DB 2; Length 910;  
Best Local Similarity 75.0%; Pred. No. 37;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CNSRLQLR 8  
|||:::  
Db 771 CNSRLQIK 778

RESULT 2

S50367

hypothetical protein YLR281c - yeast (*Saccharomyces cerevisiae*)

N;Alternate names: hypothetical protein L8003.11

C;Species: *Saccharomyces cerevisiae*

C;Date: 13-Jan-1995 #sequence\_revision 20-Feb-1995 #text\_change 19-Apr-2002

C;Accession: S50367

R;Pauley, A.

submitted to the EMBL Data Library, November 1994

A;Description: The sequence of *S. cerevisiae* cosmid 8003.

A;Reference number: S50366

A;Accession: S50367

A;Molecule type: DNA

A;Residues: 1-155 <PAU>

A;Cross-references: EMBL:U17243; NID:g596030; PIDN:AAB67327.1; PID:g596041;  
GSPDB:GN00012; MIPS:YLR281c

C;Genetics:

A;Gene: MIPS:YLR281c

A;Cross-references: SGD:S0004271

A;Map position: 12R

Query Match 70.6%; Score 36; DB 2; Length 155;  
Best Local Similarity 75.0%; Pred. No. 13;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CNSRLQLR 8  
|||::|||  
Db 56 CNSKVQLR 63

RESULT 3

T44224

hypothetical protein B7 [imported] - human herpesvirus 6 (strain Z29)

C;Species: human herpesvirus 6

A;Variety: strain Z29

C;Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 02-Jun-2000

C;Accession: T44224

R;Dominguez, G.; Dambaugh, T.R.; Stamey, F.R.; Dewhurst, S.; Inoue, N.; Pellett, P.E.

J. Virol. 73, 8040-8052, 1999

A;Title: Human herpesvirus 6B genome sequence: coding content and comparison with human herpesvirus 6A.

A;Reference number: Z22734; MUID:99412318; PMID:10482553

A;Accession: T44224

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-75 <DOM>

A;Cross-references: EMBL:AF157706; PIDN:AAB06362.1  
A;Experimental source: strain Z29; variant B  
C;Genetics:  
A;Note: B7

Query Match 68.6%; Score 35; DB 2; Length 75;  
Best Local Similarity 55.6%; Pred. No. 11;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
|:|:|:|  
Db 24 CSSRFSIRC 32

#### RESULT 4

T22010

hypothetical protein F40D4.13 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C;Accession: T22010

R;Matthews, L.

submitted to the EMBL Data Library, November 1996

A;Reference number: Z19502

A;Accession: T22010

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-340 <WIL>

A;Cross-references: EMBL:Z81536; PIDN:CAB04361.1; GSPDB:GN00023; CESP:F40D4.13

A;Experimental source: clone F40D4

C;Genetics:

A;Gene: CESP:F40D4.13

A;Map position: 5

A;Introns: 93/1; 263/3

Query Match 68.6%; Score 35; DB 2; Length 340;  
Best Local Similarity 55.6%; Pred. No. 38;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
|:|:|:|  
Db 313 CHSKVQLNC 321

#### RESULT 5

S62511

probable peptide methionine sulfoxide reductase - fission yeast  
(*Schizosaccharomyces pombe*)

C;Species: *Schizosaccharomyces pombe*

C;Date: 12-Feb-1998 #sequence\_revision 20-Feb-1998 #text\_change 10-Dec-1999

C;Accession: T38506; S62511

R;Jones, L.; Murphy, L.; McNeil, A.; Simpson, I.; Harris, D.; Barrell, B.G.;  
Rajandream, M.A.; Walsh, S.V.

submitted to the EMBL Data Library, October 1995

A;Reference number: Z21798

A;Accession: T38506

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-187 <JO2>  
A;Cross-references: EMBL:Z66525; NID:g1044926; PIDN:CAA91427.1; PID:g1044931;  
GSPDB:GN00066; SPDB:SPAC29E6.05c  
A;Experimental source: strain 972h-; cosmid c29E6  
C;Genetics:  
A;Gene: SPDB:SPAC29E6.05c  
A;Map position: 1  
C;Superfamily: peptide methionine sulfoxide reductase

Query Match 66.7%; Score 34; DB 2; Length 187;  
Best Local Similarity 44.4%; Pred. No. 35;  
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
|:|: ::|  
Db 159 CSSRMNIKC 167

RESULT 6

S34220

hypothetical protein - jelly fungus (*Trimorphomyces papilionaceus*)

C;Species: *Trimorphomyces papilionaceus*

C;Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 06-Jan-1995

C;Accession: S34220

R;Hong, S.G.

submitted to the EMBL Data Library, June 1993

A;Reference number: S34220

A;Accession: S34220

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-227 <HON>

A;Cross-references: EMBL:X73672

Query Match 66.7%; Score 34; DB 2; Length 227;  
Best Local Similarity 75.0%; Pred. No. 42;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 NSRLQLRC 9  
|| :|||  
Db 64 NSSMQLRC 71

RESULT 7

S27977

cuticle collagen dpy-7 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 05-Nov-1999

C;Accession: S27977; T34267

R;Johnstone, I.L.; Shafi, Y.; Barry, J.D.

EMBO J. 11, 3857-3863, 1992

A;Title: Molecular analysis of mutations in the *Caenorhabditis elegans* collagen gene dpy-7.

A;Reference number: S27977; MUID:93010980; PMID:1396579

A;Accession: S27977

A;Molecule type: DNA

A;Residues: 1-318 <JOH>

A;Cross-references: EMBL:X64435; NID:g6697; PIDN:CAA45773.1; PID:g6698

R;Wilcox, L.  
submitted to the EMBL Data Library, November 1995  
A;Description: The sequence of *C. elegans* cosmid F46C8.  
A;Reference number: Z21497  
A;Accession: T34267  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-318 <WIL>  
A;Cross-references: EMBL:U41624; PIDN:AAA83319.1; CESP:F46C8.6  
C;Genetics:  
A;Gene: dpy-7; CESP:F46C8.6  
A;Introns: 52/3  
C;Superfamily: unassigned collagens

Query Match 66.7%; Score 34; DB 2; Length 318;  
Best Local Similarity 66.7%; Pred. No. 56;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
| | : | | |  
Db 90 CTSCVQLRC 98

#### RESULT 8

G96675

hypothetical protein T23K8.9 [imported] - *Arabidopsis thaliana*

C;Species: *Arabidopsis thaliana* (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C;Accession: G96675

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.;  
Alonso, J.; Altaf, H.; Araujo, R.; Bowman, C.L.; Brooks, S.Y.; Buehler, E.;  
Chan, A.; Chao, Q.; Chen, H.; Cheuk, R.F.; Chin, C.W.; Chung, M.K.; Conn, L.;  
Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Dunn, P.; Etgu, P.;  
Feldblyum, T.V.; Feng, J.; Fong, B.; Fujii, C.Y.; Gill, J.E.; Goldsmith, A.D.;  
Haas, B.; Hansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.;  
Kim, C.J.; Koo, H.L.; Kremenetskaia, I.; Kurtz, D.B.; Kwan, A.; Lam, B.; Langin-  
Hooper, S.; Lee, A.; Lee, J.M.; Lenz, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu,  
S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, A.; Militscher, J.; Miranda,  
M.; Nguyen, M.; Nierman, W.C.; Osborne, B.I.; Pai, G.; Peterson, J.; Pham, P.K.;  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.;  
Tallon, L.J.; Tambunga, G.; Toriumi, M.J.; Town, C.D.; Utterback, T.; van Aken,  
S.; Vaysberg, M.; Vysotskaia, V.S.; Walker, M.; Wu, D.; Yu, G.; Fraser, C.M.;  
Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.

A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: G96675

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-653 <STO>

A;Cross-references: GB:AE005173; NID:g4646199; PIDN:AAD26872.1; GSPDB:GN00141

C;Genetics:

A;Gene: T23K8.9

A;Map position: 1

Query Match 66.7%; Score 34; DB 2; Length 653;  
Best Local Similarity 66.7%; Pred. No. 1e+02;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
|| | |||  
Db 214 CNFTLDLRC 222

RESULT 9

C96596

hypothetical protein T18I3.3 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C;Accession: C96596

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.;  
Alonso, J.; Altaf, H.; Araujo, R.; Bowman, C.L.; Brooks, S.Y.; Buehler, E.;  
Chan, A.; Chao, Q.; Chen, H.; Cheuk, R.F.; Chin, C.W.; Chung, M.K.; Conn, L.;  
Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Dunn, P.; Etgu, P.;  
Feldblyum, T.V.; Feng, J.; Fong, B.; Fujii, C.Y.; Gill, J.E.; Goldsmith, A.D.;  
Haas, B.; Hansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.;  
Kim, C.J.; Koo, H.L.; Kremenetskaia, I.; Kurtz, D.B.; Kwan, A.; Lam, B.; Langin-  
Hooper, S.; Lee, A.; Lee, J.M.; Lenz, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu,  
S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, A.; Militscher, J.; Miranda,  
M.; Nguyen, M.; Nierman, W.C.; Osborne, B.I.; Pai, G.; Peterson, J.; Pham, P.K.;  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.;  
Tallon, L.J.; Tambunga, G.; Toriumi, M.J.; Town, C.D.; Utterback, T.; van Aken,  
S.; Vaysberg, M.; Vysotskaia, V.S.; Walker, M.; Wu, D.; Yu, G.; Fraser, C.M.;  
Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: C96596

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-684 <STO>

A;Cross-references: GB:AE005173; NID:g11094789; PIDN:AAG29721.1; GSPDB:GN00141

C;Genetics:

A;Gene: T18I3.3

A;Map position: 1

Query Match 66.7%; Score 34; DB 2; Length 684;  
Best Local Similarity 66.7%; Pred. No. 1.1e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
||| | :||  
Db 545 CNSFLGIRC 553

RESULT 10

T14450

serine/threonine kinase (EC 2.7.1.-) BRLK - wild cabbage

C;Species: Brassica oleracea (wild cabbage)

C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Jun-2000

C;Accession: T14450  
 R;Stanchev, B.S.; Croy, R.R.D.  
 submitted to the EMBL Data Library, April 1997  
 A;Reference number: Z18094  
 A;Accession: T14450  
 A;Status: preliminary; translated from GB/EMBL/DDBJ  
 A;Molecule type: DNA  
 A;Residues: 1-850 <STA>  
 A;Cross-references: EMBL:Y12531  
 A;Experimental source: strain S29  
 C;Genetics:  
 A;Gene: BRLK  
 A;Introns: 467/1; 545/3; 616/1; 695/2; 744/3  
 C;Superfamily: S-receptor kinase; protein kinase homology; S-locus-specific glycoprotein homology  
 C;Keywords: ATP; phosphotransferase; serine/threonine-specific protein kinase; signal transduction  
 F;527-806/Domain: protein kinase homology <KIN>

Query Match 66.7%; Score 34; DB 2; Length 850;  
 Best Local Similarity 66.7%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
 || | |||  
 Db 766 CNKREALRC 774

# RESULT 11

T01519  
 hypothetical protein T10M13.17.1 - Arabidopsis thaliana  
 C;Species: Arabidopsis thaliana (mouse-ear cress)  
 C;Date: 19-Feb-1999 #sequence\_revision 19-Feb-1999 #text\_change 24-Mar-1999  
 C;Accession: T01519  
 R;Johnson, A.F.; de la Bastide, M.; Lodhi, M.; Hoffman, J.; Hasegawa, A.; Gnoj, L.; Gottesman, T.; Granat, S.; Hameed, A.; Kaplan, N.; Schutz, K.; Shohdy, N.; van Keuren, K.; Parnell, L.; Dedhia, N.; Martienssen, R.; McCombie, W.  
 submitted to the EMBL Data Library, May 1997  
 A;Description: The sequence of the Arabidopsis thaliana T10M13 BAC.  
 A;Reference number: Z14346  
 A;Accession: T01519  
 A;Status: translated from GB/EMBL/DDBJ  
 A;Molecule type: DNA  
 A;Residues: 1-989 <JOH>  
 A;Cross-references: EMBL:AF001308; NID:g2104523; PID:g3912931  
 A;Experimental source: cultivar Columbia  
 C;Genetics:  
 A;Map position: 4S  
 A;Introns: 31/3  
 A;Note: T10M13.17.1

Query Match 66.7%; Score 34; DB 2; Length 989;  
 Best Local Similarity 66.7%; Pred. No. 1.5e+02;  
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
 || | |||

Db

251 CNFTLDLRC 259

RESULT 12

VFIHB2

genome polyprotein - avian infectious bronchitis virus (strain Beaudette)

N;Alternate names: F2 protein

N;Contains: RNA-directed RNA polymerase (EC 2.7.7.48)

C;Species: avian infectious bronchitis virus, IBV

C;Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 11-Jun-1999

C;Accession: B33094

R;Boursnell, M.E.G.; Brown, T.D.K.; Foulds, I.J.; Green, P.F.; Tomley, F.M.; Binns, M.M.

J. Gen. Virol. 68, 57-77, 1987

A;Title: Completion of the sequence of the genome of the coronavirus avian infectious bronchitis virus.

A;Reference number: A33094; MUID:87111468; PMID:3027249

A;Accession: B33094

A;Molecule type: genomic RNA

A;Residues: 1-2652 <BOU>

A;Cross-references: GB:M94356; GB:M29496; NID:g331170; PIDN:AAA46224.1; PID:g331173

C;Superfamily: infectious bronchitis virus RNA-directed RNA polymerase

C;Keywords: glycoprotein; nucleotidyltransferase; RNA biosynthesis

F;69,543,711,726,977,1004,1240,1304,1382,1666,1795,1891,2057,2286,2317,2483,2550,2640/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 66.7%; Score 34; DB 1; Length 2652;  
Best Local Similarity 66.7%; Pred. No. 3.4e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9

|||: |||

Db 899 CNSQTILRC 907

RESULT 13

S71371

gibberellin-regulated protein GASA5 precursor - Arabidopsis thaliana

N;Alternate names: GAST1 protein homolog

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 28-Oct-1996 #sequence\_revision 27-Feb-1997 #text\_change 24-Sep-1999

C;Accession: S71371

R;Bartel, B.

submitted to the EMBL Data Library, April 1996

A;Description: A new member of the GASA gene family of Arabidopsis.

A;Reference number: S71371

A;Accession: S71371

A;Molecule type: mRNA

A;Residues: 1-97 <BAR>

A;Cross-references: EMBL:U53221; NID:g1289319; PIDN:AAA98520.1; PID:g1289320

A;Note: no signal sequence given

C;Genetics:

A;Gene: GASA5

C;Superfamily: gibberellin-regulated protein GASA2

Query Match 64.7%; Score 33; DB 2; Length 97;

Best Local Similarity 55.6%; Pred. No. 31;  
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
|||: ||  
Db 39 CNSKCSYRC 47

RESULT 14

T19366

hypothetical protein C17G1.6 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Dec-2002

C;Accession: T19366

R;White, S.

submitted to the EMBL Data Library, August 1996

A;Reference number: Z19114

A;Accession: T19366

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-736 <WIL>

A;Cross-references: EMBL:Z78415; PIDN:CAB01675.1; GSPDB:GN00028; CESP:C17G1.6

A;Experimental source: clone C17G1

C;Genetics:

A;Gene: CESP:C17G1.6

A;Map position: X

A;Introns: 23/3; 55/3; 108/1; 198/3; 234/3; 252/1; 309/1; 348/1; 379/3; 416/1;

458/1; 563/3; 612/2; 669/1; 687/2; 716/3

C;Superfamily: metalloproteinase hch-1; astacin homology

Query Match 64.7%; Score 33; DB 2; Length 736;  
Best Local Similarity 66.7%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
|| || || ||  
Db 635 CNGRLLLP 643

RESULT 15

T16840

hypothetical protein T10E10.4 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

C;Accession: T16840

R;Geisel, C.

submitted to the EMBL Data Library, October 1995

A;Description: The sequence of *C. elegans* cosmid T10E10.

A;Reference number: Z18588

A;Accession: T16840

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-1101 <GEI>

A;Cross-references: EMBL:U39644; NID:g1049339; PID:g1049343; PIDN:AAA80360.1;

CESP:T10E10.4

A;Experimental source: strain Bristol N2

C;Genetics:



A;Gene: CESP:T10E10.4

A;Introns: 93/2; 152/2; 191/3; 209/2; 283/3; 303/1; 399/3; 421/1; 440/1; 465/1;  
547/3; 765/3; 802/1; 839/1; 975/1; 1011/2; 1060/1

Query Match 64.7%; Score 33; DB 2; Length 1101;  
Best Local Similarity 55.6%; Pred. No. 2.5e+02;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
|| :||: |  
Db 487 CNQQLQMCC 495

Search completed: November 13, 2003, 09:52:55  
Job time : 10.375 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:31:40 ; Search time 5.15625 Seconds  
(without alignments)  
82.083 Million cell updates/sec

Title: US-09-228-866-5  
Perfect score: 51  
Sequence: 1 CNSRLQLRC 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result	No.	Score	% Query Match	Length	DB	ID	Description
	1	37	72.5	910	1	YO68_CAEEL	P34607 caenorhabdi
	2	37	72.5	7073	1	R1AB_CVHSA	P59641 h replicase

3	35	68.6	433	1	AS14_MOUSE	Q8vhs7	mus musculu
4	35	68.6	1288	1	CTA3_HUMAN	Q9bz76	homo sapien
5	34	66.7	302	1	AS14_HUMAN	Q8wxk2	homo sapien
6	34	66.7	318	1	CCD7_CAEEL	P34688	caenorhabdi
7	34	66.7	530	1	RAG2_BRARE	O13034	brachydanio
8	34	66.7	533	1	RAG2_ONCMY	Q91193	oncorhynchu
9	34	66.7	2652	1	RRPB_IBVB	P26314	avian infec
10	33	64.7	270	1	KITM_MOUSE	Q9r088	mus musculu
11	33	64.7	411	1	HEMZ_XENLA	O57478	xenopus lae
12	33	64.7	445	1	HGD_MOUSE	O09173	mus musculu
13	32	62.7	111	1	IF1A_METKA	Q8txz3	methanopyru
14	32	62.7	173	1	POP5_YEAST	P28005	saccharomyc
15	32	62.7	386	1	NUCM_TRYBB	P21301	trypanosoma
16	32	62.7	427	1	DHI2_SHEEP	P50168	ovis aries
17	32	62.7	482	1	ODP2_YEAST	P12695	saccharomyc
18	32	62.7	2731	1	RRPB_CVMJH	P29982	murine coro
19	32	62.7	2733	1	RRPB_CVMA5	P16342	murine coro
20	31.5	61.8	479	1	VGLC_HSV2G	P03173	herpes simp
21	31.5	61.8	480	1	VGLC_HSV23	P06475	herpes simp
22	31.5	61.8	480	1	VGLC_HSV2H	Q89730	herpes simp
23	31	60.8	50	1	HSP1_PONPY	P35310	pongo pygma
24	31	60.8	402	1	ELAD_ECOLI	Q47013	escherichia
25	31	60.8	415	1	YZ07_METJA	Q60269	methanococc
26	31	60.8	511	1	VGLC_HSV11	P10228	herpes simp
27	31	60.8	511	1	VGLC_HSV1K	P28986	herpes simp
28	31	60.8	522	1	IKAR_ONCMY	O13089	oncorhynchu
29	31	60.8	569	1	FHR5_HUMAN	Q9bxr6	homo sapien
30	31	60.8	590	1	OAM_ASCSU	Q01456	ascaris suu
31	31	60.8	760	1	SM4A_MOUSE	Q62178	mus musculu
32	31	60.8	1091	1	DIA_DROME	P48608	drosophila
33	31	60.8	1241	1	KPB1_MOUSE	P18826	mus musculu
34	31	60.8	1242	1	KPB1_RAT	Q64649	rattus norv
35	31	60.8	1378	1	RON_MOUSE	Q62190	mus musculu
36	31	60.8	1400	1	RON_HUMAN	Q04912	homo sapien
37	30	58.8	91	1	YL88_ARCFU	O28095	archaeoglob
38	30	58.8	194	1	YCEF_ECO57	P58626	escherichia
39	30	58.8	194	1	YCEF_ECOLI	P27244	escherichia
40	30	58.8	238	1	Y647_HAEIN	Q57424	haemophilus
41	30	58.8	305	1	RPO4_VACCC	P21087	vaccinia vi
42	30	58.8	305	1	RPO4_VACCV	P24757	vaccinia vi
43	30	58.8	305	1	RPO4_VARV	P33812	variola vir
44	30	58.8	367	1	TRMU_NEIMA	Q9jtt9	neisseria m
45	30	58.8	367	1	TRMU_NEIMB	Q9jyj6	neisseria m

#### ALIGNMENTS

##### RESULT 1

YO68\_CAEEL

ID YO68\_CAEEL STANDARD; PRT; 910 AA.

AC P34607;

DT 01-FEB-1994 (Rel. 28, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Hypothetical protein ZK1098.8 in chromosome III.  
GN ZK1098.8.

```

OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA   Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA   Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA   Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Sims M., Smaldon N., Smith A., Smith M., Sonnhammer E., Staden R.,
RA   Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA   Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA   Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
CC   -!- SIMILARITY: TO RIBONUCLEASE D.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; Z22176; CAA80137.1; -.
DR   PIR; S40930; S40930.
DR   WormPep; ZK1098.8; CE00370.
DR   InterPro; IPR002562; 3_5_exonuclease.
DR   Pfam; PF01612; 3_5_exonuclease; 1.
DR   SMART; SM00474; 35EXOc; 1.
KW   Hypothetical protein.
SQ   SEQUENCE 910 AA; 105569 MW; 5512D15423517FCD CRC64;

```

```

Query Match          72.5%;  Score 37;  DB 1;  Length 910;
Best Local Similarity 75.0%;  Pred. No. 11;
Matches      6;  Conservative 2;  Mismatches 0;  Indels 0;  Gaps 0;

```

```

Qy      1 CNSRLQLR 8
        |||||::
Db      771 CNSRLQIK 778

```

## RESULT 2

### R1AB\_CVHSA

```

ID   R1AB_CVHSA      STANDARD;          PRT;   7073 AA.
AC   P59641;
DT   15-SEP-2003 (Rel. 42, Created)
DT   15-SEP-2003 (Rel. 42, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)

```

DE Replicase polyprotein 1ab (pp1ab) (ORF1AB) [Includes: Replicase  
 DE polyprotein 1a (pp1a) (ORF1A)] [Contains: Leader protein; p65 homolog;  
 DE Papain-like proteinase (EC 3.4.24.-) (NSP1); 3C-like proteinase  
 DE (EC 3.4.24.-) (3CL-PRO) (NSP2); HD2 (NSP3); NSP4; NSP5; NSP6; Growth  
 DE factor-like (NSP7); RNA-directed RNA polymerase (EC 2.7.7.48) (RdRp)  
 DE (NSP9); Helicase (Hel) (NSP10); NSP11; NSP12; NSP13].  
 OS Human coronavirus (strain SARS) (HCoV-SARS).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
 OC Coronaviridae; Coronavirus.  
 OX NCBI\_TaxID=227859;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Isolate Urbani;  
 RA Bellini W.J., Campagnoli R.P., Icenogle J.P., Monroe S.S., Nix W.A.,  
 RA Oberste M.S., Pallansch M.A., Rota P.A.;  
 RL Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Isolate Tor2;  
 RA Marra M., Jones S.J.M., Holt R.;  
 RT "The complete genome of the SARS associated coronavirus.";  
 RL Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Isolate CUHK-W1;  
 RA Tsui S.K.W., Lo D.Y.M., Tam J.S., Fung K.P., Chim S.S.C., Au C.C.,  
 RA Chan A.H., Wan A.W.K., Au K.W., Chan C.W., Kou C.Y.C., Lam H.M.,  
 RA Lam W.Y., Lau S.K., Lau Y.L., Lau Y.M., Law S.L., Law T.W., Li M.L.Y.,  
 RA Tse C.H., Wong C.H., Yiu W.H., Lee C.Y., Chan A.K.C., Chiu R.W.K.,  
 RA Ng E.K.O., Tong Y.K., Chan P.K.S., Au-Yeung C., Cheung J.K.L., Chu I.,  
 RA Hung E.C.W., Waye M.M.Y.;  
 RT "DNA sequence of a human coronavirus (CUHK-W1) from a patient with  
 RT severe acute respiratory syndrome (SARS) in Hong Kong.";  
 RL Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Isolate HKU-39849;  
 RA Leung F.C., Zeng F., Chan C.W.M., Chan C.M.Y., Chen J., Chow K.Y.C.,  
 RA Hon C.C.C., Hui R.K.H., Li J., Li V.Y.Y., Wang Y.Y., Peiris J.S.M.,  
 RA Poon L.L.M.;  
 RL Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.  
 RN [5]  
 RP SEQUENCE OF 4993-5127 FROM N.A.  
 RC STRAIN=Isolate Vietnam;  
 RA Emery S., Erdman D., Peret T., Ksiazek T.;  
 RL Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.  
 RN [6]  
 RP SEQUENCE OF 4993-5136 FROM N.A.  
 RC STRAIN=Isolate Taiwan;  
 RA Lin J.-H., Chiu S.-C., Yang J.-Y., Wang S.-F., Chen H.-Y.;  
 RT "Detection of a novel human coronavirus in a severe acute respiratory  
 RT syndrome patient in Taiwan.";  
 RL Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.  
 CC -!- FUNCTION: The replicase polyprotein of coronaviruses is a  
 CC multifunctional protein: it contains the activities necessary for  
 CC the transcription of negative stranded RNA, leader RNA, subgenomic  
 CC mRNAs and progeny virion RNA as well as proteinases responsible

CC for the cleavage of the polyprotein into functional products (By  
 CC similarity).  
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC {RNA}(N).  
 CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins  
 CC (By similarity).  
 CC -!- MISCELLANEOUS: This protein is translated as a 1A-1B polyprotein  
 CC by a ribosomal frameshifting mechanism (By similarity).  
 CC -!- SIMILARITY: Contains 1 peptidase family C16 domain.  
 CC -!- SIMILARITY: Contains 1 peptidase family C30 domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AY278741; AAP13442.1; -.  
 DR EMBL; AY278741; AAP13439.1; -.  
 DR EMBL; AY278741; AAP13440.1; ALT\_SEQ.  
 DR EMBL; AY274119; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; AY278554; AAP13566.1; -.  
 DR EMBL; AY278554; AAP13575.1; -.  
 DR EMBL; AY278491; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; AY269391; AAP04003.1; -.  
 DR EMBL; AY268049; AAP04587.1; -.  
 DR InterPro; IPR002589; Alpp.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSvir.  
 DR InterPro; IPR002877; FtsJ.  
 DR Pfam; PF01661; Alpp; 1.  
 DR Pfam; PF01728; FtsJ; 1.  
 DR SMART; SM00506; Alpp; 1.  
 KW Polyprotein; Transferase; RNA-directed RNA polymerase; Thiol protease;  
 KW Hydrolase; Helicase; ATP-binding.  
 FT DOMAIN 1 179 LEADER PROTEIN (POTENTIAL).  
 FT DOMAIN 180 818 P65 HOMOLOG (POTENTIAL).  
 FT DOMAIN ? ? PAPAIN-LIKE PROTEINASE (POTENTIAL).  
 FT DOMAIN 3240 3547 3C-LIKE PROTEINASE (POTENTIAL).  
 FT DOMAIN 3548 3836 HD2/NSP3 (POTENTIAL).  
 FT DOMAIN 3837 3919 NSP4 (POTENTIAL).  
 FT DOMAIN 3920 4117 NSP5 (POTENTIAL).  
 FT DOMAIN 4118 4229 NSP6 (POTENTIAL).  
 FT DOMAIN 4230 4369 GROWTH FACTOR-LIKE (POTENTIAL).  
 FT DOMAIN 4370 5301 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).  
 FT DOMAIN 5302 5902 HELICASE (POTENTIAL).  
 FT DOMAIN 5903 6429 NSP11 (POTENTIAL).  
 FT DOMAIN 6430 6775 NSP12 (POTENTIAL).  
 FT DOMAIN 6776 7073 NSP13 (POTENTIAL).  
 FT ACT\_SITE 1909 1909 POTENTIAL.  
 FT NP\_BIND 5583 5590 ATP (POTENTIAL).  
 FT DOMAIN 930 933 POLY-GLU.  
 FT DOMAIN 937 942 POLY-GLU.  
 FT DOMAIN 974 979 POLY-GLU.  
 FT DOMAIN 2210 2213 POLY-LEU.

FT	DOMAIN	3766	3769	POLY-CYS.
FT	VARIANT	2552	2552	V -> A (in isolates Tor2, CUHK-W1 and
FT				HKU-39849).
FT	VARIANT	2556	2556	D -> N (in isolate HKU-39849).
FT	VARIANT	2708	2708	S -> T (in isolate HKU-39849).
FT	VARIANT	2718	2718	R -> T (in isolate HKU-39849).
FT	VARIANT	3047	3047	V -> A (in isolate CUHK-W1).
FT	VARIANT	3072	3072	V -> A (in isolate CUHK-W1).
FT	VARIANT	4379	4382	RVCG -> GFAV (in ORF1A).
FT	VARIANT	5131	5131	A -> G (in isolate Taiwan).
FT	VARIANT	5134	5135	CY -> VL (in isolate Taiwan).
FT	VARIANT	5767	5767	D -> E (in isolate CUHK-W1).
FT	VARIANT	6778	6778	Q -> R (in isolate Tor2).
FT	VARIANT	6883	6883	D -> Y (in isolate Tor2).
SQ	SEQUENCE	7073	AA; 790270	MW; A91B3CE920E69D4C CRC64;

Query Match 72.5%; Score 37; DB 1; Length 7073;  
 Best Local Similarity 66.7%; Pred. No. 1e+02;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
 |||: |||  
 Db 5309 CNSQTSLRC 5317

# RESULT 3

## AS14\_MOUSE

ID AS14\_MOUSE STANDARD; PRT; 433 AA.  
 AC Q8VHS7;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Ankyrin repeat and SOCS box containing protein 14 (ASB-14).  
 GN ASB14.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kile B.T., Nicola N.A.;  
 RT "SOCS box proteins."  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.  
 CC -!- SIMILARITY: Contains 9 ANK repeats.  
 CC -!- SIMILARITY: Contains 1 SOCS box domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF403042; AAL57361.1; -.  
 DR MGD; MGI:2655107; Asb14.  
 DR InterPro; IPR002110; ANK.

DR InterPro; IPR001496; SOCS.  
 DR Pfam; PF00023; ank; 8.  
 DR SMART; SM00248; ANK; 8.  
 DR PROSITE; PS50088; ANK\_REPEAT; 6.  
 DR PROSITE; PS50297; ANK\_REP\_REGION; 1.  
 DR PROSITE; PS50225; SOCS; 1.  
 KW ANK repeat; Repeat.  
 FT REPEAT 1 14 ANK 1.  
 FT REPEAT 18 47 ANK 2.  
 FT REPEAT 51 80 ANK 3.  
 FT REPEAT 94 123 ANK 4.  
 FT REPEAT 127 156 ANK 5.  
 FT REPEAT 159 188 ANK 6.  
 FT REPEAT 201 230 ANK 7.  
 FT REPEAT 231 260 ANK 8.  
 FT REPEAT 262 295 ANK 9.  
 FT DOMAIN 367 422 SOCS BOX.  
 SQ SEQUENCE 433 AA; 48317 MW; 6BCAD1AC2B2BB080 CRC64;

Query Match 68.6%; Score 35; DB 1; Length 433;  
 Best Local Similarity 66.7%; Pred. No. 12;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
 | ||:|  
 Db 391 CMGRLRLRC 399

#### RESULT 4

##### CTA3\_HUMAN

ID CTA3\_HUMAN STANDARD; PRT; 1288 AA.  
 AC Q9BZ76; Q9C0E9;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Contactin associated protein-like 3 precursor (Cell recognition molecule Caspr3).  
 GN CNTNAP3 OR CASPR3 OR KIAA1714.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=Brain;  
 RX MEDLINE=22088824; PubMed=12093160;  
 RA Spiegel I., Salomon D., Erne B., Schaeren-Wiemers N., Peles E.;  
 RT "Caspr3 and Caspr4, two novel members of the Caspr family are  
 RT expressed in the nervous system and interact with PDZ domains."  
 RL Mol. Cell. Neurosci. 20:283-297(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE=Brain;  
 RA Nagase T., Kikuno R., Yamakawa H., Ohara O.;  
 RL Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE OF 71-1288 FROM N.A. (ISOFORM 2).

```

RC  TISSUE=Brain;
RX  MEDLINE=21082932; PubMed=11214970;
RA  Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;
RT  "Prediction of the coding sequences of unidentified human genes. XIX.
RT  The complete sequences of 100 new cDNA clones from brain which code
RT  for large proteins in vitro.";
RL  DNA Res. 7:347-355(2000).
CC  -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential). Isoform
CC      2 seems to be secreted.
CC  -!- ALTERNATIVE PRODUCTS:
CC      Event=Alternative splicing; Named isoforms=2;
CC      Name=1;
CC      IsoId=Q9BZ76-1; Sequence=Displayed;
CC      Name=2;
CC      IsoId=Q9BZ76-2; Sequence=VSP_003535, VSP_003536;
CC      Note=No experimental confirmation available;
CC  -!- SIMILARITY: Contains 1 F5/8 type C domain.
CC  -!- SIMILARITY: Contains 4 laminin G-like domains.
CC  -!- SIMILARITY: Contains 2 EGF-like domains.
CC  -!- SIMILARITY: BELONGS TO THE NEUREXIN FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AF333769; AAG52889.2; -.
DR  EMBL; AB051501; BAB21805.2; ALT_INIT.
DR  HSSP; P12259; 1CZT.
DR  GO; GO:0016021; C:integral to membrane; NAS.
DR  GO; GO:0005194; F:cell adhesion molecule activity; NAS.
DR  GO; GO:0008037; P:cell recognition; NAS.
DR  InterPro; IPR006209; EGF_like.
DR  InterPro; IPR000421; FA58_C.
DR  InterPro; IPR002181; Fibrinogen_C.
DR  InterPro; IPR006210; IEGF.
DR  InterPro; IPR001791; Laminin_G.
DR  Pfam; PF00008; EGF; 2.
DR  Pfam; PF00754; F5_F8_type_C; 1.
DR  Pfam; PF00054; laminin_G; 3.
DR  SMART; SM00181; EGF; 2.
DR  SMART; SM00231; FA58C; 1.
DR  SMART; SM00186; FBG; 1.
DR  SMART; SM00282; LamG; 4.
DR  PROSITE; PS00022; EGF_1; FALSE_NEG.
DR  PROSITE; PS01186; EGF_2; FALSE_NEG.
DR  PROSITE; PS01285; FA58C_1; 1.
DR  PROSITE; PS01286; FA58C_2; 1.
DR  PROSITE; PS50022; FA58C_3; 1.
DR  PROSITE; PS00514; FIBRIN_AG_C_DOMAIN; FALSE_NEG.
DR  PROSITE; PS50025; LAM_G_DOMAIN; 4.
KW  Glycoprotein; Cell adhesion; Signal; Transmembrane; Repeat;
KW  Alternative splicing.
FT  SIGNAL      1      25      POTENTIAL.

```



FT	CHAIN	26	1288	CONTACTIN ASSOCIATED PROTEIN-LIKE 3.
FT	DOMAIN	26	1245	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1246	1266	POTENTIAL.
FT	DOMAIN	1267	1288	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	42	48	POLY-SER.
FT	DOMAIN	31	177	F5/8 TYPE C.
FT	DOMAIN	183	364	LAMININ G-LIKE 1.
FT	DOMAIN	370	545	LAMININ G-LIKE 2.
FT	DOMAIN	551	583	EGF-LIKE 1.
FT	DOMAIN	793	958	LAMININ G-LIKE 3.
FT	DOMAIN	962	996	EGF-LIKE 2.
FT	DOMAIN	1015	1203	LAMININ G-LIKE 4.
FT	CARBOHYD	285	285	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	359	359	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	441	441	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	497	497	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	623	623	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	706	706	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1023	1023	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1073	1073	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1120	1120	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPPLIC	1120	1127	NQSTKKQV -> IPQMOKSN (in isoform 2).
FT				/FTId=VSP_003535.
FT	VARSPPLIC	1128	1288	Missing (in isoform 2).
FT				/FTId=VSP_003536.
FT	CONFLICT	21	21	R -> S (IN REF. 2).
FT	CONFLICT	33	33	S -> A (IN REF. 2).
FT	CONFLICT	89	89	I -> M (IN REF. 2).
FT	CONFLICT	714	714	G -> V (IN REF. 2).
FT	CONFLICT	769	771	TGQ -> AGR (IN REF. 2).
FT	CONFLICT	777	777	D -> A (IN REF. 2).
FT	CONFLICT	845	845	R -> H (IN REF. 2).
SQ	SEQUENCE	1288 AA;	140878 MW;	C31C3564032787D1 CRC64;

Query Match 68.6%; Score 35; DB 1; Length 1288;  
 Best Local Similarity 66.7%; Pred. No. 40;  
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
 | || |||  
 Db 676 CEQRLALRC 684

# RESULT 5

AS14\_HUMAN

ID AS14\_HUMAN STANDARD; PRT; 302 AA.

AC Q8WXX2;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Ankyrin repeat and SOCS box containing protein 14 (ASB-14).

GN ASB14.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.  
 RA Kile B.T., Nicola N.A.;  
 RT "SOCS box proteins.";  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.  
 CC -!- SIMILARITY: Contains 4 ANK repeats.  
 CC -!- SIMILARITY: Contains 1 SOCS box domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF403032; AAL57351.1; -.  
 DR Genew; HGNC:19766; ASB14.  
 DR InterPro; IPR002110; ANK.  
 DR InterPro; IPR001496; SOCS.  
 DR Pfam; PF00023; ank; 3.  
 DR SMART; SM00248; ANK; 4.  
 DR PROSITE; PS50088; ANK\_REPEAT; 2.  
 DR PROSITE; PS50297; ANK\_REP\_REGION; 1.  
 DR PROSITE; PS50225; SOCS; 1.  
 KW ANK repeat; Repeat.  
 FT REPEAT 28 57 ANK 1.  
 FT REPEAT 70 99 ANK 2.  
 FT REPEAT 100 129 ANK 3.  
 FT REPEAT 131 164 ANK 4.  
 FT DOMAIN 236 291 SOCS BOX.  
 SQ SEQUENCE 302 AA; 34562 MW; 0B8C6E7219E9EF7B CRC64;

Query Match 66.7%; Score 34; DB 1; Length 302;  
 Best Local Similarity 66.7%; Pred. No. 13;  
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CNSRLQLRC 9  
 | ||||  
 Db 260 CMGRLHLRC 268

#### RESULT 6

##### CCD7\_CAEEL

ID CCD7\_CAEEL STANDARD; PRT; 318 AA.  
 AC P34688;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cuticle collagen dpy-7 precursor.  
 GN DPY-7 OR F46C8.6.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93010980; PubMed=1396579;

RA Johnstone I.L., Shafi Y., Barry J.D.;  
 RT "Molecular analysis of mutations in the *Caenorhabditis elegans*  
 RT collagen gene dpy-7.";  
 RL EMBO J. 11:3857-3863(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Wilcox L.;  
 RL Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.  
 CC -!- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE  
 CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A  
 CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.  
 CC -!- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE  
 CC CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-  
 CC LINKS.  
 CC -!- DISEASE: MUTATIONS IN DPY-7 AFFECTS THE BODY SHAPE.  
 CC -!- SIMILARITY: BELONGS TO THE CUTICULAR COLLAGEN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X64435; CAA45773.1; -.  
 DR EMBL; U41624; AAF99944.1; -.  
 DR PIR; S27977; S27977.  
 DR WormPep; F46C8.6; CE04580.  
 DR InterPro; IPR002486; Col\_cuticle\_N.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF01484; Col\_cuticle\_N; 1.  
 DR Pfam; PF01391; Collagen; 3.  
 KW Cuticle; Connective tissue; Repeat; Multigene family; Collagen;  
 KW Signal.  
 FT SIGNAL 1 ? POTENTIAL.  
 FT CHAIN ? 318 CUTICLE COLLAGEN DPY-7.  
 FT DOMAIN 101 130 TRIPLE-HELICAL REGION.  
 FT DOMAIN 147 206 TRIPLE-HELICAL REGION.  
 FT DOMAIN 209 235 TRIPLE-HELICAL REGION.  
 FT DOMAIN 240 278 TRIPLE-HELICAL REGION.  
 FT VARIANT 101 101 G -> R (IN DPY7(SC27)).  
 FT VARIANT 156 156 G -> R (IN DPY7(E88)).  
 FT VARIANT 189 189 G -> Y (IN DPY7(E1234)).  
 FT VARIANT 201 201 G -> R (IN DPY7(M38)).  
 SQ SEQUENCE 318 AA; 31629 MW; 4EA66DA5FDC5737C CRC64;

Query Match 66.7%; Score 34; DB 1; Length 318;  
 Best Local Similarity 66.7%; Pred. No. 14;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
 | | : | | |  
 Db 90 CTSCVQLRC 98

RESULT 7

RAG2\_BRARE

ID RAG2\_BRARE STANDARD; PRT; 530 AA.  
 AC O13034;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE V(D)J recombination activating protein 2 (RAG-2).  
 GN RAG2 OR RAG-2.  
 OS Brachydanio rerio (Zebrafish) (Danio rerio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 OX NCBI\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND DEVELOPMENTAL STAGE.  
 RC TISSUE=Larva;  
 RX MEDLINE=97246732; PubMed=9089097;  
 RA Willett C.E., Cherry J.J., Steiner L.A.;  
 RT "Characterization and expression of the recombination activating genes  
 RT (rag1 and rag2) of zebrafish.";  
 RL Immunogenetics 45:394-404(1997).  
 RN [2]  
 RP DEVELOPMENTAL STAGE.  
 RX MEDLINE=97223529; PubMed=9070331;  
 RA Willett C.E., Zapata A.G., Hopkins N., Steiner L.A.;  
 RT "Expression of zebrafish rag genes during early development identifies  
 RT the thymus.";  
 RL Dev. Biol. 182:331-341(1997).

CC -!- FUNCTION: During lymphocyte development, the genes encoding  
 CC immunoglobulins and T cell receptors are assembled from variable  
 CC (V), diversity (D), and joining (J) gene segments. This  
 CC combinatorial process, known as V(D)J recombination, allows the  
 CC generation of an enormous range of binding specificities from a  
 CC limited amount of genetic information. The RAG1/RAG2 complex  
 CC initiates this process by binding to the conserved recombination  
 CC signal sequences (RSS) and introducing a double-strand break  
 CC between the RSS and the adjacent coding segment. These breaks are  
 CC generated in two steps, nicking of one strand (hydrolysis),  
 CC followed by hairpin formation (transesterification). RAG1/2 has  
 CC also been shown to function as a transposase in vitro, and to  
 CC possess RSS-independent endonuclease activity (end processing) and  
 CC hairpin opening. RAG1 alone can bind to RSS but stable, efficient  
 CC binding requires RAG2. All known catalytic activities require the  
 CC presence of both proteins (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- DEVELOPMENTAL STAGE: First detected in the thymus during day 4 of  
 CC development. Expression then increases in the thymus for at least  
 CC three weeks.  
 CC -!- SIMILARITY: BELONGS TO THE RAG2 FAMILY.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch).

CC -----  
DR EMBL; U71094; AAC60366.1; -.  
DR ZFIN; ZDB-GENE-990415-235; rag2.  
DR InterPro; IPR004321; RAG2.  
DR Pfam; PF03089; RAG2; 1.  
KW Hydrolase; Endonuclease; Nuclear protein; DNA-binding;  
KW DNA recombination.  
FT DOMAIN 352 412 ASP/GLU-RICH (ACIDIC).  
SQ SEQUENCE 530 AA; 59173 MW; 2E96CD0C3B9F1417 CRC64;

Query Match 66.7%; Score 34; DB 1; Length 530;  
Best Local Similarity 55.6%; Pred. No. 24;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
|| :: |||  
Db 116 CNRKVTLRRC 124

#### RESULT 8

##### RAG2\_ONCMY

ID RAG2\_ONCMY STANDARD; PRT; 533 AA.  
AC Q91193;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE V(D)J recombination activating protein 2 (RAG-2).  
GN RAG2.  
OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.  
OX NCBI\_TaxID=8022;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Shasta;  
RX MEDLINE=96270000; PubMed=8662087;  
RA Hansen J.D., Kaattari S.L.;  
RT "The recombination activating gene 2 (RAG2) of the rainbow trout  
RT Oncorhynchus mykiss.";  
RL Immunogenetics 44:203-211(1996).  
CC -!- FUNCTION: During lymphocyte development, the genes encoding  
CC immunoglobulins and T cell receptors are assembled from variable  
CC (V), diversity (D), and joining (J) gene segments. This  
CC combinatorial process, known as V(D)J recombination, allows the  
CC generation of an enormous range of binding specificities from a  
CC limited amount of genetic information. The RAG1/RAG2 complex  
CC initiates this process by binding to the conserved recombination  
CC signal sequences (RSS) and introducing a double-strand break  
CC between the RSS and the adjacent coding segment. These breaks are  
CC generated in two steps, nicking of one strand (hydrolysis),  
CC followed by hairpin formation (transesterification). RAG1/2 has  
CC also been shown to function as a transposase in vitro, and to  
CC possess RSS-independent endonuclease activity (end processing) and  
CC hairpin opening. RAG1 alone can bind to RSS but stable, efficient  
CC binding requires RAG2. All known catalytic activities require the

CC presence of both proteins (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- SIMILARITY: BELONGS TO THE RAG2 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U31670; AAB18138.1; -.  
 DR EMBL; U25146; AAA65927.1; -.  
 DR InterPro; IPR004321; RAG2.  
 DR Pfam; PF03089; RAG2; 1.  
 KW Hydrolase; Endonuclease; Nuclear protein; DNA-binding;  
 KW DNA recombination.  
 SQ SEQUENCE 533 AA; 59410 MW; 18AE5F4B79096D83 CRC64;

Query Match 66.7%; Score 34; DB 1; Length 533;  
 Best Local Similarity 55.6%; Pred. No. 24;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
 || :: |||  
 Db 116 CNRKVTLC 124

# RESULT 9

## RRPB\_IBVB

ID RRPB\_IBVB STANDARD; PRT; 2652 AA.  
 AC P26314;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE RNA-directed RNA polymerase (ORF1B) (EC 2.7.7.48).  
 GN F2.  
 OS Avian infectious bronchitis virus (strain Beaudette) (IBV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
 OC Coronaviridae; Coronavirus.  
 OX NCBI\_TaxID=11122;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87111468; PubMed=3027249;  
 RA Bourns M.E.G., Brown T.D.K., Foulds I.J., Green P.F., Tomley F.M.,  
 RA Binns M.M.;  
 RT "Completion of the sequence of the genome of the coronavirus avian  
 RT infectious bronchitis virus."  
 RL J. Gen. Virol. 68:57-77(1987).  
 CC -!- FUNCTION: THE RNA DEPENDENT RNA POLYMERASE OF CORONAVIRUSES IS  
 CC A MULTIFUNCTIONAL PROTEIN: IT CONTAINS THE ACTIVITIES NECESSARY  
 CC FOR THE TRANSCRIPTION OF NEGATIVE STRANDED RNA, LEADER RNA,  
 CC SUBGENOMIC MRNAS AND PROGENY VIRION RNA.  
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC {RNA} (N).  
 CC -!- MISCELLANEOUS: THIS PROTEIN IS EXPRESSED BY AN EFFICIENT RIBOSOMAL

CC FRAMESHIFTING MECHANISM. RIBOSOMAL FRAMESHIFTING IS AN ELEGANT  
CC MECHANISM FOR REGULATING THE SYNTHESIS OF SEVERAL PROTEINS IN A  
CC WELL BALANCED MANNER.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; M94356; AAA46224.1; -.  
DR EMBL; M95169; AAA70234.1; -.  
DR PIR; B33094; VFIHB2.  
DR InterPro; IPR003593; AAA\_ATPase.  
DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro; IPR007094; RNA\_pol\_PSVir.  
DR InterPro; IPR000606; Viral\_helicase1.  
DR Pfam; PF01443; Viral\_helicase1; 1.  
DR SMART; SM00382; AAA; 1.  
KW Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.  
FT NP\_BIND 1173 1180 ATP (BY SIMILARITY).  
SQ SEQUENCE 2652 AA; 300617 MW; F5D7DBFD09D1E29D CRC64;

Query Match 66.7%; Score 34; DB 1; Length 2652;  
Best Local Similarity 66.7%; Pred. No. 1.4e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CNSRLQLRC 9  
|||: |||  
Db 899 CNSQTILRC 907

#### RESULT 10

##### KITM\_MOUSE

ID KITM\_MOUSE STANDARD; PRT; 270 AA.  
AC Q9R088;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Thymidine kinase 2, mitochondrial precursor (EC 2.7.1.21) (Mt-TK).  
GN TK2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20035846; PubMed=10571069;  
RA Wettin K., Johansson M., Zheng X., Zhu C., Karlsson A.;  
RT "Cloning of mouse mitochondrial thymidine kinase 2 cDNA.";  
RL FEBS Lett. 460:103-106(1999).  
RN [2]  
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
RC STRAIN=C57BL/6; TISSUE=Brain;  
RX MEDLINE=20480069; PubMed=11023833;

RA Wang L., Eriksson S.;  
 RT "Cloning and characterisation of full length mouse thymidine kinase 2:  
 RT the N -terminal sequence directs import of the precursor protein into  
 RT mitochondria.";  
 RL Biochem. J. 351:469-476(2000).  
 CC -!- FUNCTION: DEOXYRIBONUCLEOSIDE KINASE THAT PHOSPHORYLATES  
 CC THYMIDINE, DEOXYCYTIDINE, AND DEOXYURIDINE. ALSO PHOSPHORYLATES  
 CC ANTI-VIRAL AND ANTI-CANCER NUCLEOSIDE ANALOGS.  
 CC -!- CATALYTIC ACTIVITY: ATP + thymidine = ADP + thymidine 5'-  
 CC phosphate.  
 CC -!- SUBUNIT: Homodimer.  
 CC -!- SUBCELLULAR LOCATION: Mitochondrial.  
 CC -!- TISSUE SPECIFICITY: FOUND IN MOST TISSUES; HIGHLY EXPRESSED IN  
 CC LIVER.  
 CC -!- SIMILARITY: BELONGS TO THE DCK/DGK FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF105217; AAF08104.1; -.  
 DR EMBL; AJ249341; CAC07190.2; -.  
 DR MGD; MGI:1913266; Tk2.  
 DR GO; GO:0005739; C:mitochondrion; IDA.  
 DR InterPro; IPR002624; dNK.  
 DR Pfam; PF01712; dNK; 1.  
 KW Transferase; Kinase; DNA synthesis; ATP-binding; Mitochondrion;  
 KW Transit peptide.  
 FT TRANSIT 1 38 MITOCHONDRION (POTENTIAL).  
 FT CHAIN 39 270 THYMIDINE KINASE 2.  
 FT NP\_BIND 62 69 ATP (POTENTIAL).  
 FT CONFLICT 14 14 P -> L (IN REF. 1).  
 FT CONFLICT 23 23 G -> R (IN REF. 1).  
 FT CONFLICT 155 155 G -> S (IN REF. 1).  
 FT CONFLICT 269 270 GP -> WTLGLSDLQDSARNSPARARCHGPRA (IN REF.  
 FT 1).  
 SQ SEQUENCE 270 AA; 31209MW; 886F5B80D2C3EFE2 CRC64;

Query Match 64.7%; Score 33; DB 1; Length 270;  
 Best Local Similarity 55.6%; Pred. No. 18;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CNSRLQLRC 9  
 | ||::||  
 Db 194 CYQRLKMRC 202

RESULT 11  
 HEMZ\_XENLA  
 ID HEMZ\_XENLA STANDARD; PRT; 411 AA.  
 AC O57478;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)



DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Ferrochelatase, mitochondrial precursor (EC 4.99.1.1) (Protoheme  
 DE ferro-lyase) (Heme synthetase).  
 GN FECH.  
 OS *Xenopus laevis* (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
 OC Xenopodinae; *Xenopus*.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RX MEDLINE=99027642; PubMed=9808757;  
 RA Day A.L., Parsons B.M., Dailey H.A.;  
 RT "Cloning and characterization of Gallus and *Xenopus* ferrochelatases:  
 RT presence of the [2Fe-2S] cluster in nonmammalian ferrochelatase.";   
 RL Arch. Biochem. Biophys. 359:160-169(1998).  
 CC -!- FUNCTION: CATALYZES THE FERROUS INSERTION INTO PROTOPORPHYRIN IX.  
 CC -!- CATALYTIC ACTIVITY: Protoporphyrin + Fe(2+) = protoheme + 2 H(+).  
 CC -!- COFACTOR: BINDS 1 2FE-2S CLUSTER.  
 CC -!- PATHWAY: Protoheme biosynthesis; last step.  
 CC -!- SUBUNIT: Monomer (By similarity).  
 CC -!- SUBCELLULAR LOCATION: BOUND TO THE MITOCHONDRIAL INNER MEMBRANE IN  
 CC EUKARYOTIC CELLS WITH ITS ACTIVE SITE ON THE MATRIX SIDE OF THE  
 CC MEMBRANE (BY SIMILARITY).  
 CC -!- SIMILARITY: Belongs to the ferrochelatase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF036617; AAB94626.1; -.  
 DR HSSP; P32396; 1AK1.  
 DR InterPro; IPR001015; Ferrochelatase.  
 DR Pfam; PF00762; Ferrochelatase; 1.  
 DR ProDom; PD002792; Ferrochelatase; 1.  
 DR TIGRFAMS; TIGR00109; hemH; 1.  
 DR PROSITE; PS00534; FERROCHELATASE; 1.  
 KW Porphyrin biosynthesis; Heme biosynthesis; Lyase; Mitochondrion;  
 KW Transit peptide; Metal-binding; Iron-sulfur; Iron; 2Fe-2S.  
 FT TRANSIT 1 41 MITOCHONDRION (POTENTIAL).  
 FT CHAIN 42 411 FERROCHELATASE.  
 FT METAL 183 183 IRON-SULFUR (2FE-2S).  
 FT METAL 390 390 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 FT METAL 393 393 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 FT METAL 398 398 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 FT ACT\_SITE 217 217 BY SIMILARITY.  
 FT ACT\_SITE 370 370 BY SIMILARITY.  
 SQ SEQUENCE 411 AA; 46039 MW; 010A1C422697A2B3 CRC64;

Query Match 64.7%; Score 33; DB 1; Length 411;  
 Best Local Similarity 55.6%; Pred. No. 28;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CNSRLQLRC 9  
|: :| |||  
Db 382 CSKQLSLRC 390

RESULT 12

HGD\_MOUSE

ID HGD\_MOUSE STANDARD; PRT; 445 AA.  
AC O09173;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Homogentisate 1,2-dioxygenase (EC 1.13.11.5) (Homogentisicase)  
DE (Homogentisate oxygenase) (Homogentisic acid oxidase).  
GN HGD OR HGO OR AKU.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN=C57BL/6 X CBA; TISSUE=Liver;  
RX MEDLINE=97222472; PubMed=9069115;  
RA Schmidt S.R., Gehrig A., Koehler M.R., Schmid M., Mueller C.R.,  
RA Kress W.;  
RT "Cloning of the homogentisate 1,2-dioxygenase gene, the key enzyme of  
RT alkaptonuria in mouse.";  
RL Mamm. Genome 8:168-171(1997).  
RN [2]  
RP CHARACTERIZATION.  
RC TISSUE=Liver;  
RX MEDLINE=95220372; PubMed=7705358;  
RA Schmidt S.R., Muller C.R., Kress W.;  
RT "Murine liver homogentisate 1,2-dioxygenase. Purification to  
RT homogeneity and novel biochemical properties.";  
RL Eur. J. Biochem. 228:425-430(1995).  
CC -!- CATALYTIC ACTIVITY: Homogentisate + O(2) = 4-maleylacetoacetate.  
CC -!- COFACTOR: IRON.  
CC -!- PATHWAY: Catabolism of tyrosine; third step.  
CC -!- PATHWAY: Catabolism of phenylalanine; fourth step.  
CC -!- SUBUNIT: Homotrimer (Probable).  
CC -!- DISEASE: DEFECTS IN HGD ARE THE CAUSE OF ALKAPTONURIA (AKU), AN  
CC AUTOSOMAL RECESSIVE ERROR OF METABOLISM. AKU IS CHARACTERIZED BY  
CC AN INCREASE IN THE LEVEL OF HOMOGENTISIC ACID.  
CC -!- SIMILARITY: Belongs to the homogentisate dioxygenase family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U58988; AAC53224.1; -.  
DR HSSP; Q93099; 1EYB.  
DR MGD; MGI:96078; Hgd.

DR InterPro; IPR005708; HmgA.  
 DR Pfam; PF04209; HgmA; 1.  
 DR TIGRFAMs; TIGR01015; hmgA; 1.  
 KW Oxidoreductase; Dioxygenase; Iron; Phenylalanine catabolism;  
 KW Tyrosine catabolism.  
 FT METAL 335 335 IRON (BY SIMILARITY).  
 FT METAL 341 341 IRON (BY SIMILARITY).  
 FT METAL 371 371 IRON (BY SIMILARITY).  
 SQ SEQUENCE 445 AA; 49990 MW; C7CBBCFD3764B93F CRC64;

Query Match 64.7%; Score 33; DB 1; Length 445;  
 Best Local Similarity 55.6%; Pred. No. 31;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
 ||| :: ||  
 Db 138 CNSSMENRC 146

# RESULT 13

## IF1A\_METKA

ID IF1A\_METKA STANDARD; PRT; 111 AA.  
 AC Q8TXZ3;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Translation initiation factor 1A (aIF-1A).  
 GN EIF1A OR MK0515.  
 OS Methanopyrus kandleri.  
 OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;  
 OC Methanopyrus.  
 OX NCBI\_TaxID=2320;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AV19 / DSM 6324 / JCM 9639;  
 RX MEDLINE=21927647; PubMed=11930014;  
 RA Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,  
 RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,  
 RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,  
 RA Malykh A.G., Koonin E.V., Kozyavkin S.A.;  
 RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19  
 RT and monophyly of archaeal methanogens.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).  
 CC -!- FUNCTION: Seems to be required for maximal rate of protein  
 CC biosynthesis. Enhances ribosome dissociation into subunits and  
 CC stabilizes the binding of the initiator Met-tRNA(I) to 40 S  
 CC ribosomal subunits (By similarity).  
 CC -!- SIMILARITY: Belongs to the eIF-1A family.  
 CC -!- SIMILARITY: Contains 1 S1-like domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC -----
DR   EMBL; AE010345; AAM01730.1; -.
DR   HAMAP; MF_00216; -; 1.
DR   InterPro; IPR006196; S1_IF1.
DR   InterPro; IPR001253; TIF_eIF-1A.
DR   Pfam; PF01176; eIF-1a; 1.
DR   ProDom; PD005579; TIF_eIF-1A; 1.
DR   SMART; SM00652; eIF1a; 1.
DR   TIGRFAMs; TIGR00523; eIF-1A; 1.
DR   PROSITE; PS01262; IF1A; 1.
DR   PROSITE; PS50832; S1_IF1_TYPE; 1.
KW   Initiation factor; Protein biosynthesis; Complete proteome.
FT   DOMAIN      11      83      S1-LIKE.
SQ   SEQUENCE    111 AA;  13083 MW;  93F67811814199A8 CRC64;

```

```

Query Match          62.7%;  Score 32;  DB 1;  Length 111;
Best Local Similarity 62.5%;  Pred. No. 11;
Matches      5;  Conservative      2;  Mismatches      1;  Indels      0;  Gaps      0;

```

```

Qy      2 NSRLQLRC 9
        | | : | : | |
Db      31 NDRVQVRC 38

```

#### RESULT 14

##### POP5\_YEAST

```

ID   POP5_YEAST      STANDARD;          PRT;   173 AA.
AC   P28005;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Ribonucleases P/MRP protein subunit POP5 (EC 3.1.26.5) (RNases P/MRP
DE   19.6 kDa subunit) (RNA processing protein POP5).
GN   POP5 OR YAL033W OR FUN53.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92260538; PubMed=1583694;
RA   Harris S.D., Cheng J., Pugh T.A., Pringle J.R.;
RT   "Molecular analysis of Saccharomyces cerevisiae chromosome I. On the
RT   number of genes and the identification of essential genes using
RT   temperature-sensitive-lethal mutations.";
RL   J. Mol. Biol. 225:53-65(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=95249563; PubMed=7731988;
RA   Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,
RA   Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,
RA   Storms R.K.;
RT   "The nucleotide sequence of chromosome I from Saccharomyces
RT   cerevisiae.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).
CC   -!- FUNCTION: COMPONENT OF RIBONUCLEASE P, A PROTEIN COMPLEX THAT

```

CC GENERATES MATURE TRNA MOLECULES BY CLEAVING THEIR 5' ENDS.  
 CC ALSO A COMPONENT OF RNASE MRP.  
 CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage of RNA, removing 5'-  
 CC extra-nucleotide from tRNA precursor.  
 CC -!- SUBUNIT: COMPONENT OF NUCLEAR RNASE P AND RNASE MRP RNASE P  
 CC RIBONUCLEOPROTEINS. RNASE P CONSISTS OF A RNA MOIETY AND AT LEAST  
 CC 8 PROTEIN SUBUNITS; POP1, POP3, POP4, POP5, POP6, POP7, POP8 AND  
 CC RPP1.  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X62577; CAA44457.1; -.  
 DR EMBL; U12980; AAC04999.1; -.  
 DR PIR; S23411; S23411.  
 DR SGD; S0000031; POP5.  
 DR GO; GO:0000172; C:ribonuclease mitochondrial RNA processing c. . .; IDA.  
 DR GO; GO:0005655; C:ribonuclease P complex; IDA.  
 DR GO; GO:0000171; F:ribonuclease MRP activity; IDA.  
 DR GO; GO:0004526; F:ribonuclease P activity; IDA.  
 DR InterPro; IPR002759; RNase\_P\_related.  
 DR Pfam; PF01900; RNase\_P\_Rpp14; 1.  
 DR ProDom; PD012772; RNase\_P\_related; 1.  
 KW Hydrolase; Nuclear protein; tRNA processing.  
 SQ SEQUENCE 173 AA; 19573 MW; 918193631BD790DD CRC64;

Query Match 62.7%; Score 32; DB 1; Length 173;  
 Best Local Similarity 75.0%; Pred. No. 17;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CNSRLQLR 8  
 ||| |||:  
 Db 68 CNSLLQLK 75

# RESULT 15

## NUCM\_TRYBB

ID NUCM\_TRYBB STANDARD; PRT; 386 AA.  
 AC P21301;  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE NADH-ubiquinone oxidoreductase 49 kDa subunit homolog (EC 1.6.5.3)  
 DE (NADH dehydrogenase subunit 7 homolog).  
 GN NAD7 OR MURF 3.  
 OS Trypanosoma brucei brucei.  
 OG Mitochondrion.  
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_TaxID=5702;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RX MEDLINE=90367122; PubMed=2393904;  
 RA Koslowsky D.J., Bhat G., Jayarama Perrollaz A.L., Feagin J.E.,  
 RA Stuart K.;  
 RT "The MURF3 gene of *T. brucei* contains multiple domains of extensive  
 RT editing and is homologous to a subunit of NADH dehydrogenase.";  
 RL Cell 62:901-911(1990).  
 CC -!- FUNCTION: TRANSFER OF ELECTRONS FROM NADH TO THE RESPIRATORY  
 CC CHAIN. THE IMMEDIATE ELECTRON ACCEPTOR FOR THE ENZYME IS BELIEVED  
 CC TO BE UBIQUINONE. COMPONENT OF THE IRON-SULFUR (IP) FRAGMENT OF  
 CC THE ENZYME.  
 CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.  
 CC -!- SUBCELLULAR LOCATION: Mitochondrial.  
 CC -!- SIMILARITY: BELONGS TO THE COMPLEX I 49 kDa SUBUNIT FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M55645; -; NOT\_ANNOTATED\_CDS.  
 DR PIR; A35693; A35693.  
 DR InterPro; IPR001135; Oxidored\_49kDa.  
 DR Pfam; PF00346; complex1\_49Kd; 1.  
 DR PROSITE; PS00535; COMPLEX1\_49K; 1.  
 KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Kinetoplast.  
 SQ SEQUENCE 386 AA; 45098 MW; 448F5D52DC572071 CRC64;

Query Match 62.7%; Score 32; DB 1; Length 386;  
 Best Local Similarity 85.7%; Pred. No. 42;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 SRLQLRC 9  
 |||:|||  
 Db 343 SRLRLRC 349

Search completed: November 13, 2003, 09:46:34  
 Job time : 6.15625 secs

GenCore version 5.1.6  
 Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:31:40 ; Search time 23.7188 Seconds  
 (without alignments)  
 97.917 Million cell updates/sec

Title: US-09-228-866-5  
 Perfect score: 51  
 Sequence: 1 CNSRLQLRC 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL\_23:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	%		DB	ID	Description
		Query	Match Length			
1	37	72.5	175	4	Q8WVH1	Q8wvh1 homo sapien
2	37	72.5	236	4	Q8J028	Q8j028 homo sapien
3	37	72.5	245	11	Q8R217	Q8r217 mus musculu
4	37	72.5	332	13	Q98U07	Q98u07 pseudotylos
5	37	72.5	332	13	Q98U08	Q98u08 platybelone
6	37	72.5	333	13	Q9DF04	Q9df04 strongylura
7	37	72.5	333	13	Q9DF15	Q9df15 platybelone
8	37	72.5	333	13	Q9DF08	Q9df08 strongylura
9	37	72.5	333	13	Q9DF10	Q9df10 potamorrhap
10	37	72.5	333	13	Q9DF14	Q9df14 potamorrhap
11	37	72.5	333	13	Q9DF01	Q9df01 belonion ap
12	37	72.5	333	13	Q9DD82	Q9dd82 potamorrhap
13	37	72.5	333	13	Q9DD51	Q9dd51 pseudotylos
14	37	72.5	333	13	Q9DD50	Q9dd50 belonion di
15	37	72.5	333	13	Q9DF03	Q9df03 strongylura
16	37	72.5	333	13	Q9DF16	Q9df16 strongylura

17	37	72.5	333	13	Q9DF12	Q9df12 strongylura
18	37	72.5	333	13	Q9DD64	Q9dd64 strongylura
19	37	72.5	333	13	Q9DD35	Q9dd35 strongylura
20	37	72.5	333	13	Q9DF13	Q9df13 potamorrhap
21	37	72.5	333	13	Q9DF05	Q9df05 strongylura
22	37	72.5	333	13	Q9DF02	Q9df02 strongylura
23	37	72.5	333	13	Q9DF09	Q9df09 strongylura
24	37	72.5	333	13	Q9DF17	Q9df17 strongylura
25	37	72.5	333	13	Q9DF00	Q9df00 strongylura
26	37	72.5	333	13	Q9DF06	Q9df06 strongylura
27	37	72.5	333	13	Q9DF11	Q9df11 xenentodon
28	37	72.5	333	13	Q9DF07	Q9df07 scomberesox
29	37	72.5	482	5	Q95TI4	Q95ti4 drosophila
30	37	72.5	482	5	Q9VP72	Q9vp72 drosophila
31	37	72.5	611	13	Q9IBF6	Q9ibf6 xenopus lae
32	37	72.5	611	13	Q9PTI0	Q9pti0 xenopus lae
33	37	72.5	1086	5	Q9N976	Q9n976 leishmania
34	36	70.6	142	10	Q94DJ7	Q94dj7 oryza sativ
35	36	70.6	148	5	Q9W4U3	Q9w4u3 drosophila
36	36	70.6	155	3	Q05863	Q05863 saccharomyc
37	36	70.6	1087	13	Q91778	Q91778 xenopus lae
38	36	70.6	4138	5	Q8I1Y3	Q8ily3 plasmodium
39	35	68.6	75	12	Q69066	Q69066 human herpe
40	35	68.6	167	11	Q8BMJ9	Q8bmj9 mus musculus
41	35	68.6	340	5	Q9XV33	Q9xv33 caenorhabdi
42	35	68.6	604	4	Q96NU0	Q96nu0 homo sapien
43	35	68.6	627	16	Q8Y006	Q8y006 ralstonia s
44	35	68.6	653	10	Q8W5H2	Q8w5h2 oryza sativ
45	35	68.6	745	4	Q96MJ5	Q96mj5 homo sapien

# ALIGNMENTS

## RESULT 1

Q8WVH1

ID Q8WVH1 PRELIMINARY; PRT; 175 AA.  
AC Q8WVH1;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Hypothetical protein (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RA Strausberg R.;  
RL Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; BC018019; AAH18019.1; -.  
DR InterPro; IPR000967; Znf\_NFX1.  
DR Pfam; PF01422; zf-NF-X1; 1.  
KW Hypothetical protein.  
FT NON\_TER 1 1  
SQ SEQUENCE 175 AA; 20439 MW; 072F35C835DC122B CRC64;



Query Match 72.5%; Score 37; DB 4; Length 175;  
Best Local Similarity 55.6%; Pred. No. 5.2;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
|| ::|||  
Db 56 CNQKVKLRC 64

RESULT 2

Q8J028

ID Q8J028 PRELIMINARY; PRT; 236 AA.  
AC Q8J028;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Human ovarian zinc finger protein.  
GN HOZFP.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Ovary;  
RA Guo J.H., Yu L.;  
RL Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AY134856; AAN08626.1; -.  
SQ SEQUENCE 236 AA; 27114 MW; D49108CB443A299F CRC64;

Query Match 72.5%; Score 37; DB 4; Length 236;  
Best Local Similarity 55.6%; Pred. No. 6.7;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
|| ::|||  
Db 117 CNQKVKLRC 125

RESULT 3

Q8R217

ID Q8R217 PRELIMINARY; PRT; 245 AA.  
AC Q8R217;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Hypothetical 28.0 kDa protein.  
GN AW538212.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Colon;  
RA Strausberg R.;

RL Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; BC022652; AAH22652.1; -.  
DR MGD; MGI:2141210; AW538212.  
DR InterPro; IPR000967; Znf\_NFX1.  
DR Pfam; PF01422; zf-NF-X1; 2.  
DR SMART; SM00438; ZnF\_NFX; 3.  
KW Hypothetical protein.  
SQ SEQUENCE 245 AA; 27958 MW; 22B986095B2137A7 CRC64;

Query Match 72.5%; Score 37; DB 11; Length 245;  
Best Local Similarity 55.6%; Pred. No. 7;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
|| ::|||  
Db 126 CNQKVKLRC 134

#### RESULT 4

Q98U07

ID Q98U07 PRELIMINARY; PRT; 332 AA.  
AC Q98U07;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Recombination-activating protein 2 (Fragment).  
GN RAG2.  
OS Pseudotylosurus angusticeps.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;  
OC Beloniformes; Belonidae; Pseudotylosurus.  
OX NCBI\_TaxID=106211;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=N28b;  
RA Lovejoy N.R., Collette B.B.;  
RT "Phylogenetic relationships of New World needlefishes (Teleostei:  
RT Belonidae) and the biogeography of transitions between marine and  
RT freshwater habitats."  
RL Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AF306476; AAG23200.2; -.  
DR InterPro; IPR004321; RAG2.  
DR Pfam; PF03089; RAG2; 1.  
FT NON\_TER 1 1  
FT NON\_TER 332 332  
SQ SEQUENCE 332 AA; 36738 MW; 53F77F52C6B6698A CRC64;

Query Match 72.5%; Score 37; DB 13; Length 332;  
Best Local Similarity 66.7%; Pred. No. 9.1;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
|| :| |||  
Db 79 CNRKLTLC 87

# RESULT 5

Q98U08

ID Q98U08 PRELIMINARY; PRT; 332 AA.  
AC Q98U08;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Recombination-activating protein 2 (Fragment).  
GN RAG2.  
OS Platybelone argalus (Keeltail needlefish).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;  
OC Beloniformes; Belonidae; Platybelone.  
OX NCBI\_TaxID=129059;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=N12a;  
RA Lovejoy N.R., Collette B.B.;  
RT "Phylogenetic relationships of New World needlefishes (Teleostei:  
RT Belonidae) and the biogeography of transitions between marine and  
RT freshwater habitats.";  
RL Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AF306464; AAG23188.2; -.  
DR InterPro; IPR004321; RAG2.  
DR Pfam; PF03089; RAG2; 1.  
FT NON\_TER 1 1  
FT NON\_TER 332 332  
SQ SEQUENCE 332 AA; 36757 MW; 58F379B877990CAC CRC64;

Query Match 72.5%; Score 37; DB 13; Length 332;  
Best Local Similarity 66.7%; Pred. No. 9.1;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
|| :| |||  
Db 79 CNRKLTLC 87

# RESULT 6

Q9DF04

ID Q9DF04 PRELIMINARY; PRT; 333 AA.  
AC Q9DF04;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Recombination-activating protein 2 (Fragment).  
GN RAG2.  
OS Strongylura senegalensis.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;  
OC Beloniformes; Belonidae; Strongylura.  
OX NCBI\_TaxID=106208;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=N39a;

RA Lovejoy N.R., Collette B.B.;  
 RT "Phylogenetic relationships of New World needlefishes (Teleostei:  
 RT Belonidae) and the biogeography of transitions between marine and  
 RT freshwater habitats.";  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF306484; AAG23208.1; -.  
 DR InterPro; IPR004321; RAG2.  
 DR Pfam; PF03089; RAG2; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 333 333  
 SQ SEQUENCE 333 AA; 36727 MW; EAC22D977D09F0DE CRC64;  
  
 Query Match 72.5%; Score 37; DB 13; Length 333;  
 Best Local Similarity 66.7%; Pred. No. 9.2;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CNSRLQLRC 9  
 || :| |||  
 Db 79 CNRKLTLRC 87

# RESULT 7

Q9DF15

ID Q9DF15 PRELIMINARY; PRT; 333 AA.  
 AC Q9DF15;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Recombination-activating protein 2 (Fragment).  
 GN RAG2.  
 OS Platybelone argalus (Keeltail needlefish).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;  
 OC Beloniformes; Belonidae; Platybelone.  
 OX NCBI\_TaxID=129059;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=N12b;  
 RA Lovejoy N.R., Collette B.B.;  
 RT "Phylogenetic relationships of New World needlefishes (Teleostei:  
 RT Belonidae) and the biogeography of transitions between marine and  
 RT freshwater habitats.";  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF306465; AAG23189.1; -.  
 DR InterPro; IPR004321; RAG2.  
 DR Pfam; PF03089; RAG2; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 333 333  
 SQ SEQUENCE 333 AA; 36822 MW; 9048DBD8E9CDBE02 CRC64;  
  
 Query Match 72.5%; Score 37; DB 13; Length 333;  
 Best Local Similarity 66.7%; Pred. No. 9.2;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CNSRLQLRC 9  
 || :| |||

Db 79 CNRKLTLRC 87

RESULT 8

Q9DF08

ID Q9DF08 PRELIMINARY; PRT; 333 AA.  
AC Q9DF08;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Recombination-activating protein 2 (Fragment).  
GN RAG2.  
OS Strongylura hubbsi.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;  
OC Beloniformes; Belonidae; Strongylura.  
OX NCBI\_TaxID=129064;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=N30b;  
RA Lovejoy N.R., Collette B.B.;  
RT "Phylogenetic relationships of New World needlefishes (Teleostei:  
RT Belonidae) and the biogeography of transitions between marine and  
RT freshwater habitats.";  
RL Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AF306480; AAG23204.1; -.  
DR InterPro; IPR004321; RAG2.  
DR Pfam; PF03089; RAG2; 1.  
FT NON\_TER 1 1  
FT NON\_TER 333 333  
SQ SEQUENCE 333 AA; 36828 MW; C4D6FFFD5EF0E524 CRC64;

Query Match 72.5%; Score 37; DB 13; Length 333;  
Best Local Similarity 66.7%; Pred. No. 9.2;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
|| :| |||  
Db 79 CNRKLTLRC 87

RESULT 9

Q9DF10

ID Q9DF10 PRELIMINARY; PRT; 333 AA.  
AC Q9DF10;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Recombination-activating protein 2 (Fragment).  
GN RAG2.  
OS Potamorrhaphis petersi.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;  
OC Beloniformes; Belonidae; Potamorrhaphis.  
OX NCBI\_TaxID=105858;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=N27;  
 RA Lovejoy N.R., Collette B.B.;  
 RT "Phylogenetic relationships of New World needlefishes (Teleostei:  
 RT Belonidae) and the biogeography of transitions between marine and  
 RT freshwater habitats.";  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF306474; AAG23198.1; -.  
 DR InterPro; IPR004321; RAG2.  
 DR Pfam; PF03089; RAG2; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 333 333  
 SQ SEQUENCE 333 AA; 36647 MW; CCD3DE04952A99C9 CRC64;  
  
 Query Match 72.5%; Score 37; DB 13; Length 333;  
 Best Local Similarity 66.7%; Pred. No. 9.2;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
 || :| |||  
 Db 79 CNRKLTLCRC 87

# RESULT 10

Q9DF14  
 ID Q9DF14 PRELIMINARY; PRT; 333 AA.  
 AC Q9DF14;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Recombination-activating protein 2 (Fragment).  
 GN RAG2.  
 OS Potamorrhaphis guianensis.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;  
 OC Beloniformes; Belonidae; Potamorrhaphis.  
 OX NCBI\_TaxID=105857;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=N13a;  
 RA Lovejoy N.R., Collette B.B.;  
 RT "Phylogenetic relationships of New World needlefishes (Teleostei:  
 RT Belonidae) and the biogeography of transitions between marine and  
 RT freshwater habitats.";  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF306466; AAG23190.1; -.  
 DR InterPro; IPR004321; RAG2.  
 DR Pfam; PF03089; RAG2; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 333 333  
 SQ SEQUENCE 333 AA; 36783 MW; 8FC4B19CADB9842A CRC64;  
  
 Query Match 72.5%; Score 37; DB 13; Length 333;  
 Best Local Similarity 66.7%; Pred. No. 9.2;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
 || :| |||  
 Db 79 CNRKLTLC 87

RESULT 11

Q9DF01

ID Q9DF01 PRELIMINARY; PRT; 333 AA.  
 AC Q9DF01;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Recombination-activating protein 2 (Fragment).  
 GN RAG2.  
 OS Belonion apodion.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;  
 OC Beloniformes; Belonidae; Belonion.  
 OX NCBI\_TaxID=105853;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=N55;  
 RA Lovejoy N.R., Collette B.B.;  
 RT "Phylogenetic relationships of New World needlefishes (Teleostei:  
 RT Belonidae) and the biogeography of transitions between marine and  
 RT freshwater habitats.";  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF306488; AAG23212.1; -.  
 DR InterPro; IPR004321; RAG2.  
 DR Pfam; PF03089; RAG2; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 333 333  
 SQ SEQUENCE 333 AA; 36648 MW; 967AED83BE6879D3 CRC64;

Query Match 72.5%; Score 37; DB 13; Length 333;  
 Best Local Similarity 66.7%; Pred. No. 9.2;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
 || :| |||  
 Db 79 CNRKLTLC 87

RESULT 12

Q9DD82

ID Q9DD82 PRELIMINARY; PRT; 333 AA.  
 AC Q9DD82;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Recombination-activating protein 2 (Fragment).  
 GN RAG2.  
 OS Potamorhaphis eigenmanni.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;  
 OC Beloniformes; Belonidae; Potamorhaphis.  
 OX NCBI\_TaxID=105855;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=N18, and N17;  
 RA Lovejoy N.R., Collette B.B.;  
 RT "Phylogenetic relationships of New World needlefishes (Teleostei:  
 RT Belonidae) and the biogeography of transitions between marine and  
 RT freshwater habitats.";  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF306471; AAG23195.1; -.  
 DR EMBL; AF306470; AAG23194.1; -.  
 DR InterPro; IPR004321; RAG2.  
 DR Pfam; PF03089; RAG2; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 333 333  
 SQ SEQUENCE 333 AA; 36684 MW; 0FC075EC5D110DA7 CRC64;

Query Match 72.5%; Score 37; DB 13; Length 333;  
 Best Local Similarity 66.7%; Pred. No. 9.2;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
 || :| |||  
 Db 79 CNRKLTLC 87

# RESULT 13

Q9DD51

ID Q9DD51 PRELIMINARY; PRT; 333 AA.  
 AC Q9DD51;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Recombination-activating protein 2 (Fragment).  
 GN RAG2.  
 OS Pseudotylosurus angusticeps.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;  
 OC Beloniformes; Belonidae; Pseudotylosurus.  
 OX NCBI\_TaxID=106211;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=N41, and N28a;  
 RA Lovejoy N.R., Collette B.B.;  
 RT "Phylogenetic relationships of New World needlefishes (Teleostei:  
 RT Belonidae) and the biogeography of transitions between marine and  
 RT freshwater habitats.";  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF306486; AAG23210.1; -.  
 DR EMBL; AF306475; AAG23199.1; -.  
 DR InterPro; IPR004321; RAG2.  
 DR Pfam; PF03089; RAG2; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 333 333



SQ SEQUENCE 333 AA; 36826 MW; A363F77F52C6B669 CRC64;

Query Match 72.5%; Score 37; DB 13; Length 333;  
Best Local Similarity 66.7%; Pred. No. 9.2;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
|| :| |||  
Db 79 CNRKLTLC 87

#### RESULT 14

Q9DD50

ID Q9DD50 PRELIMINARY; PRT; 333 AA.  
AC Q9DD50;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Recombination-activating protein 2 (Fragment).  
GN RAG2.  
OS Belonion dibranchodon.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;  
OC Beloniformes; Belonidae; Belonion.  
OX NCBI\_TaxID=105856;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=N14b, and N14a;  
RA Lovejoy N.R., Collette B.B.;  
RT "Phylogenetic relationships of New World needlefishes (Teleostei:  
RT Belonidae) and the biogeography of transitions between marine and  
RT freshwater habitats.";  
RL Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AF306469; AAG23193.1; -.  
DR EMBL; AF306468; AAG23192.1; -.  
DR InterPro; IPR004321; RAG2.  
DR Pfam; PF03089; RAG2; 1.  
FT NON\_TER 1 1  
FT NON\_TER 333 333  
SQ SEQUENCE 333 AA; 36611 MW; ABC1F33B243E4422 CRC64;

Query Match 72.5%; Score 37; DB 13; Length 333;  
Best Local Similarity 66.7%; Pred. No. 9.2;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
|| :| |||  
Db 79 CNRKLTLC 87

#### RESULT 15

Q9DF03

ID Q9DF03 PRELIMINARY; PRT; 333 AA.  
AC Q9DF03;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Recombination-activating protein 2 (Fragment).  
 GN RAG2.  
 OS Strongylura senegalensis.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;  
 OC Beloniformes; Belonidae; Strongylura.  
 OX NCBI\_TaxID=106208;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=N39b;  
 RA Lovejoy N.R., Collette B.B.;  
 RT "Phylogenetic relationships of New World needlefishes (Teleostei:  
 RT Belonidae) and the biogeography of transitions between marine and  
 RT freshwater habitats.";  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF306485; AAG23209.1; -.  
 DR InterPro; IPR004321; RAG2.  
 DR Pfam; PF03089; RAG2; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 333 333  
 SQ SEQUENCE 333 AA; 36734 MW; EAC22D8B10BF37DE CRC64;

Query Match 72.5%; Score 37; DB 13; Length 333;  
 Best Local Similarity 66.7%; Pred. No. 9.2;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
 || :| |||  
 Db 79 CNRKLTLCRC 87

Search completed: November 13, 2003, 09:51:01  
 Job time : 24.7188 secs

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:39:50 ; Search time 10.6875 Seconds  
 (without alignments)  
 35.630 Million cell updates/sec

Title: US-09-228-866-5  
 Perfect score: 51  
 Sequence: 1 CNSRLQLRC 9

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
 1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
 2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
 3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
 4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
 5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	51	100.0	9	1	US-08-526-710-5
2	51	100.0	9	3	US-08-862-855-5
3	51	100.0	9	3	US-09-226-985-5
4	51	100.0	9	4	US-09-227-906-5
5	46	90.2	9	1	US-08-526-710-1
6	46	90.2	9	3	US-08-862-855-1
7	46	90.2	9	3	US-09-226-985-1
8	46	90.2	9	4	US-09-227-906-1
9	33	64.7	445	4	US-09-252-991A-20277
10	32	62.7	270	4	US-09-252-991A-22459
11	32	62.7	270	4	US-09-252-991A-26033

12	32	62.7	371	4	US-09-199-637A-295	Sequence 295, App
13	32	62.7	371	4	US-09-252-991A-21430	Sequence 21430, A
14	31	60.8	154	4	US-09-252-991A-31454	Sequence 31454, A
15	31	60.8	263	4	US-09-252-991A-28276	Sequence 28276, A
16	31	60.8	509	4	US-09-252-991A-22513	Sequence 22513, A
17	31	60.8	511	1	US-08-220-151-17	Sequence 17, Appl
18	31	60.8	511	1	US-08-413-118-17	Sequence 17, Appl
19	31	60.8	511	3	US-08-473-446-17	Sequence 17, Appl
20	31	60.8	1073	4	US-09-252-991A-30317	Sequence 30317, A
21	31	60.8	1388	4	US-09-252-991A-20237	Sequence 20237, A
22	30.5	59.8	220	4	US-09-252-991A-24410	Sequence 24410, A
23	30.5	59.8	363	4	US-09-252-991A-17517	Sequence 17517, A
24	30	58.8	62	4	US-09-134-001C-3739	Sequence 3739, Ap
25	30	58.8	119	4	US-09-107-532A-4666	Sequence 4666, Ap
26	30	58.8	172	4	US-09-252-991A-21771	Sequence 21771, A
27	30	58.8	209	4	US-09-252-991A-17766	Sequence 17766, A
28	30	58.8	225	2	US-08-951-871-4	Sequence 4, Appli
29	30	58.8	319	4	US-09-489-847-130	Sequence 130, App
30	30	58.8	332	4	US-09-252-991A-16753	Sequence 16753, A
31	30	58.8	341	3	US-09-008-465-1	Sequence 1, Appli
32	30	58.8	341	4	US-09-528-959-1	Sequence 1, Appli
33	30	58.8	450	4	US-09-252-991A-32284	Sequence 32284, A
34	30	58.8	549	1	US-08-325-071-61	Sequence 61, Appl
35	30	58.8	549	3	US-08-461-004A-61	Sequence 61, Appl
36	30	58.8	552	3	US-08-796-899-28	Sequence 28, Appl
37	30	58.8	620	1	US-08-325-071-65	Sequence 65, Appl
38	30	58.8	620	3	US-08-461-004A-65	Sequence 65, Appl
39	30	58.8	650	1	US-08-325-071-63	Sequence 63, Appl
40	30	58.8	650	1	US-08-325-071-67	Sequence 67, Appl
41	30	58.8	650	3	US-08-461-004A-63	Sequence 63, Appl
42	30	58.8	650	3	US-08-461-004A-67	Sequence 67, Appl
43	30	58.8	1048	4	US-09-171-699-10	Sequence 10, Appl
44	29.5	57.8	272	4	US-09-686-583B-2	Sequence 2, Appli
45	29	56.9	21	1	US-08-016-023A-5	Sequence 5, Appli

#### ALIGNMENTS

##### RESULT 1

US-08-526-710-5

; Sequence 5, Application US/08526710

; Patent No. 5622699

; GENERAL INFORMATION:

; APPLICANT: Ruoslahti, Erkki

; APPLICANT: Pasqualini, Renata

; TITLE OF INVENTION: Method of Identifying Molecules That

; TITLE OF INVENTION: Home to a Selected Organ In Vivo

; NUMBER OF SEQUENCES: 44

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Campbell and Flores

; STREET: 4370 La Jolla Village Drive, Suite 700

; CITY: San Diego

; STATE: California

; COUNTRY: United States

; ZIP: 92122

; COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/526,710
; FILING DATE: 11-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-526-710-5

```

```

Query Match          100.0%; Score 51; DB 1; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      9; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      1 CNSRLQLRC 9
        |||||
Db      1 CNSRLQLRC 9

```

## RESULT 2

```

US-08-862-855-5
; Sequence 5, Application US/08862855
; Patent No. 6068829
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/862,855
; FILING DATE:

```

```

; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2621
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-862-855-5

```

```

Query Match          100.0%; Score 51; DB 3; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 CNSRLQLRC 9
        |||||
Db      1 CNSRLQLRC 9

```

```

RESULT 3
US-09-226-985-5
; Sequence 5, Application US/09226985
; Patent No. 6296832
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Molecules That Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/226,985
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,855
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3423
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-226-985-5

```

```

Query Match          100.0%; Score 51; DB 3; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      9; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      1 CNSRLQLRC 9
        |||||
Db      1 CNSRLQLRC 9

```

#### RESULT 4

```

US-09-227-906-5
; Sequence 5, Application US/09227906
; Patent No. 6306365
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/227,906
; FILING DATE:

```

```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,855
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3424
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-227-906-5

```

```

Query Match          100.0%; Score 51; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches      9; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      1 CNSRLQLRC 9
        |||||
Db      1 CNSRLQLRC 9

```

# RESULT 5

```

US-08-526-710-1
; Sequence 1, Application US/08526710
; Patent No. 5622699
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```



```

; APPLICATION NUMBER: US/08/526,710
; FILING DATE: 11-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-526-710-1

```

```

Query Match          90.2%; Score 46; DB 1; Length 9;
Best Local Similarity 88.9%; Pred. No. 2.5e+05;
Matches      8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 CNSRLQLRC 9
        ||||| |||
Db      1 CNSRLHLRC 9

```

# RESULT 6

```

US-08-862-855-1
; Sequence 1, Application US/08862855
; Patent No. 6068829
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/862,855
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2621
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-862-855-1

```

```

Query Match          90.2%; Score 46; DB 3; Length 9;
Best Local Similarity 88.9%; Pred. No. 2.5e+05;
Matches      8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 CNSRLQLRC 9
        ||||| |||
Db      1 CNSRLHLRC 9

```

# RESULT 7

US-09-226-985-1

```

; Sequence 1, Application US/09226985
; Patent No. 6296832
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Molecules That Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/226,985
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,855
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3423
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-226-985-1

```

```

Query Match          90.2%; Score 46; DB 3; Length 9;
Best Local Similarity 88.9%; Pred. No. 2.5e+05;
Matches      8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 CNSRLQLRC 9
        ||||| |||
Db      1 CNSRLHLRC 9

```

# RESULT 8

US-09-227-906-1

```

; Sequence 1, Application US/09227906
; Patent No. 6306365
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; TITLE OF INVENTION: Method of Identifying Molecules That
; TITLE OF INVENTION: Home to a Selected Organ In Vivo
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/227,906
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,710
; FILING DATE: 11-SEP-1995
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US 08/813,273
; FILING DATE: 10-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/862,855
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3424
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-227-906-1

```

```

Query Match          90.2%; Score 46; DB 4; Length 9;
Best Local Similarity 88.9%; Pred. No. 2.5e+05;
Matches      8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 CNSRLQLRC 9
        ||||| |||
Db      1 CNSRLHLRC 9

```

```

RESULT 9
US-09-252-991A-20277
; Sequence 20277, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 20277
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20277

```

```

Query Match          64.7%; Score 33; DB 4; Length 445;
Best Local Similarity 55.6%; Pred. No. 2e+02;
Matches      5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy      1 CNSRLQLRC 9

```

Db                   |||       :||  
36 CNSSTSMRC 44

RESULT 10

US-09-252-991A-22459  
; Sequence 22459, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 22459  
; LENGTH: 270  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-22459

Query Match                   62.7%; Score 32; DB 4; Length 270;  
Best Local Similarity   44.4%; Pred. No. 1.8e+02;  
Matches   4; Conservative   3; Mismatches   2; Indels   0; Gaps   0;

Qy                   1 CNSRLQLRC 9  
                     | :| ::||  
Db                   186 CRARAEIRC 194

RESULT 11

US-09-252-991A-26033  
; Sequence 26033, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 26033  
; LENGTH: 270  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-26033

Query Match 62.7%; Score 32; DB 4; Length 270;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 RLQLRC 9  
|||  
Db 114 RLQLRC 119

RESULT 12

US-09-199-637A-295

; Sequence 295, Application US/09199637A  
; Patent No. 6355411  
; GENERAL INFORMATION:  
; APPLICANT: Ausubel, Frederick  
; APPLICANT: Goodman, Howard M.  
; APPLICANT: Rahme, Laurence G.  
; APPLICANT: Mahajan-Miklos, Shalina  
; APPLICANT: Tan, Man-Wah  
; APPLICANT: Cao, Hui  
; APPLICANT: Drenkard, Eliana  
; APPLICANT: Tsongalis, John  
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID  
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF  
; FILE REFERENCE: 00786/361002  
; CURRENT APPLICATION NUMBER: US/09/199,637A  
; CURRENT FILING DATE: 1998-11-25  
; PRIOR APPLICATION NUMBER: 60/066,517  
; PRIOR FILING DATE: 1997-11-25  
; NUMBER OF SEQ ID NOS: 437  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 295  
; LENGTH: 371  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa

US-09-199-637A-295

Query Match 62.7%; Score 32; DB 4; Length 371;  
Best Local Similarity 55.6%; Pred. No. 2.5e+02;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CNSRLQLRC 9  
|:| :|  
Db 316 CSSRAESRC 324

RESULT 13

US-09-252-991A-21430

; Sequence 21430, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

```
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21430
; LENGTH: 371
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21430
```

```
Query Match          62.7%; Score 32; DB 4; Length 371;
Best Local Similarity 55.6%; Pred. No. 2.5e+02;
Matches      5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy      1 CNSRLQLRC 9
          |:| | : ||
Db      316 CSSRAESRC 324
```

#### RESULT 14

```
US-09-252-991A-31454
; Sequence 31454, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31454
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31454
```

```
Query Match          60.8%; Score 31; DB 4; Length 154;
Best Local Similarity 55.6%; Pred. No. 1.6e+02;
Matches      5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy      1 CNSRLQLRC 9
          |:| | : ||
Db      2 CSSSLGIRC 10
```

#### RESULT 15

```
US-09-252-991A-28276
```

```

; Sequence 28276, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28276
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28276

```

```

Query Match          60.8%; Score 31; DB 4; Length 263;
Best Local Similarity 55.6%; Pred. No. 2.6e+02;
Matches      5; Conservative    2; Mismatches    2; Indels      0; Gaps      0;

```

```

Qy          1 CNSRLQLRC 9
              | :| ||:|
Db          42 CAARAQLQC 50

```

```

Search completed: November 13, 2003, 09:54:57
Job time : 11.6875 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:31:40 ; Search time 23.5521 Seconds  
(without alignments)  
47.176 Million cell updates/sec

Title: US-09-228-866-6  
Perfect score: 43  
Sequence: 1 CGVRLGC 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_19Jun03:\*

- 1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:\*
- 2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:\*
- 3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:\*
- 4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:\*
- 5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:\*
- 6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:\*
- 7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:\*
- 8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:\*
- 9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:\*
- 10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:\*
- 11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:\*
- 12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:\*
- 13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:\*
- 14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:\*
- 15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:\*
- 16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:\*
- 17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:\*
- 18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:\*
- 19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:\*
- 20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:\*
- 21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:\*
- 22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:\*
- 23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:\*
- 24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

# SUMMARIES

Result		%					
No.	Score	Query	Match	Length	DB	ID	Description
1	43	100.0		7	18	AAW13417	Brain homing pepti
2	43	100.0		7	21	AAB07392	Brain homing pepti
3	43	100.0		7	22	AAE11798	Phage peptide #6 t
4	43	100.0		7	23	AAU10709	Brain homing pepti
5	36	83.7		61	22	AAO02039	Human polypeptide
6	36	83.7		159	22	AAG67507	Amino acid sequenc
7	36	83.7		434	22	AAB48195	Drosophila mutant
8	36	83.7		438	22	ABB61858	Drosophila melanog
9	36	83.7		438	22	ABB67347	Drosophila melanog
10	36	83.7		438	22	AAB48188	Drosophila wild-ty
11	36	83.7		438	22	AAB48191	Drosophila mutant
12	36	83.7		438	22	AAB48192	Drosophila mutant
13	36	83.7		438	22	AAB48193	Drosophila mutant
14	36	83.7		438	22	AAB48194	Drosophila mutant
15	36	83.7		438	22	AAB48196	Drosophila mutant
16	36	83.7		438	22	AAB48197	Drosophila mutant
17	36	83.7		585	23	ABU05348	Pancreas-originate
18	36	83.7		662	16	AAR73595	Cotransporter prot
19	36	83.7		674	23	ABU05342	Pancreas-originate
20	36	83.7		674	23	ABU05346	Pancreas-originate
21	36	83.7		674	23	ABU05347	Pancreas-originate
22	36	83.7		678	23	ABU05343	Pancreas-originate
23	36	83.7		681	23	ABU05344	Pancreas-originate
24	36	83.7		681	23	AAO14202	Human transporter
25	36	83.7		684	24	ABJ37930	NOVX protein seque
26	36	83.7		704	24	ABJ37934	NOVX protein seque
27	36	83.7		720	23	ABP69719	Human polypeptide
28	36	83.7		742	23	AAE16778	Human transporter
29	36	83.7		743	24	ABJ37932	NOVX protein seque
30	36	83.7		752	22	ABG28100	Novel human diagno
31	36	83.7		752	23	AAE16783	Human transporter
32	36	83.7		1335	22	ABB71593	Drosophila melanog
33	36	83.7		1922	22	ABB63631	Drosophila melanog
34	35	81.4		21	22	ABG58476	Human liver peptid
35	35	81.4		21	22	ABB43076	Peptide #10582 enc
36	35	81.4		21	22	ABB26233	Protein #8232 enco
37	35	81.4		21	22	AAM63975	Human brain expres
38	35	81.4		21	22	AAM76795	Human bone marrow
39	35	81.4		21	22	AAM21004	Peptide #7438 enco
40	35	81.4		21	22	AAM36901	Peptide #10938 enc
41	35	81.4		21	23	ABG45954	Human peptide enco
42	35	81.4		232	22	AAG91073	C glutamicum prote
43	35	81.4		322	22	ABB69471	Drosophila melanog
44	35	81.4		342	21	AAB29472	Burkholderia sp. C
45	34	79.1		15	19	AAW82252	CTLA-4 immunomodul

# ALIGNMENTS

RESULT 1

AAW13417

ID AAW13417 standard; Peptide; 7 AA.

XX

AC AAW13417;

XX

DT 15-JAN-1998 (first entry)

XX

DE Brain homing peptide.

XX

KW Brain homing peptide; in vivo panning; screening; phage display;  
KW drug delivery.

XX

OS Synthetic.

XX

PN WO9710507-A1.

XX

PD 20-MAR-1997.

XX

PF 10-SEP-1996; 96WO-US14600.

XX

PR 11-SEP-1995; 95US-0526710.

PR 11-SEP-1995; 95US-0526708.

XX

PA (LJOL-) LA JOLLA CANCER RES FOUND.

XX

PI Pasqualini R, Ruoslahti E;

XX

DR WPI; 1997-202359/18.

XX

PT Obtaining compound that homes to selected organ or tissue - by in  
PT vivo panning method, specifically to identify brain, kidney,  
PT angiogenic vasculature or tumour tissue homing peptide(s)

XX

PS Claim 14; Page 67; 75pp; English.

XX

CC This synthetic peptide is a claimed example of a brain-homing  
CC peptide that was identified using a novel method for obtaining  
CC molecules that home to a selected organ or tissue. This in vivo  
CC panning method typically involves administering a phage display  
CC library to a subject, and identifying expressed peptides which  
CC home to the desired organ or tissue, e.g. brain, kidney, angiogenic  
CC vascular tissue or tumour tissue. The isolated peptides (see  
CC AAW13412-52, AAW11181-86) can be used to target e.g. drugs, toxins or  
CC labels to the selected organ/tissue (claimed) or to identify and/or  
CC isolate target molecules (claimed). The peptides can be directly  
CC identified in vivo, as compared to prior art in vitro screening  
CC methods, which require further examination to see if they maintain  
CC specificity in vivo.

XX

SQ Sequence 7 AA;

Query Match 100.0%; Score 43; DB 18; Length 7;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7

Db

|||||||  
1 CGVRLGC 7

RESULT 2

AAB07392

ID AAB07392 standard; peptide; 7 AA.

XX

AC AAB07392;

XX

DT 17-OCT-2000 (first entry)

XX

DE Brain homing peptide # 6.

XX

KW Brain; homing peptide; organ targeting; tissue targeting; mouse; cyclic.

XX

OS Mus sp.

XX

FH Key Location/Qualifiers

FT Disulfide-bond 1..7

FT /note= "Can optionally form a cyclic peptide"

XX

PN US6068829-A.

XX

PD 30-MAY-2000.

XX

PF 23-JUN-1997; 97US-0862855.

XX

PR 11-SEP-1995; 95US-0526710.

PR 10-MAR-1997; 97US-0813273.

XX

PA (BURN-) BURNHAM INST..

XX

PI Pasqualini R, Ruoslahti E;

XX

DR WPI; 2000-410850/35.

XX

PT Identifying and recovering organ homing molecules or peptides by in  
PT vivo panning comprises administering a library of diverse peptides  
PT linked to a tag which facilitates recovery of these peptides -

XX

PS Example 2; Column 17; 20pp; English.

XX

CC The present sequence is a mouse brain homing peptide. This sequence was  
CC identified by using in vivo panning to screen a library of potential  
CC organ homing molecules. The present sequence can be used to direct a  
CC moiety to a the brain tissue, by linking the moiety to the present  
CC sequence. Examples of potential moieties are drugs, toxins or a  
CC detectable label. The present sequence contains a VRL amino acid motif.

XX

SQ Sequence 7 AA;

Query Match 100.0%; Score 43; DB 21; Length 7;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy

1 CGVRLGC 7

Db                    |||||  
                      1 CGVRLGC 7

RESULT 3

AAE11798

ID    AAE11798 standard; peptide; 7 AA.

XX

AC    AAE11798;

XX

DT    18-DEC-2001    (first entry)

XX

DE    Phage peptide #6 targetted to brain.

XX

KW    Enriched library fraction; brain; kidney; tumour; panning; diagnostic;

KW    molecular medicine; drug delivery; peptidomimetic; pharmaceutical.

XX

OS    Bacteriophage.

XX

FH    Key                    Location/Qualifiers

FT    Domain                3..5

FT                         /label= VLR\_motif

XX

PN    US6296832-B1.

XX

PD    02-OCT-2001.

XX

PF    08-JAN-1999;    99US-0226985.

XX

PR    23-JUN-1997;    97US-0862855.

PR    11-SEP-1995;    95US-0526710.

PR    10-MAR-1997;    97US-0813273.

XX

PA    (BURN-) BURNHAM INST.

XX

PI    Ruoslahti E,    Pasqualini R;

XX

DR    WPI; 2001-610691/70.

XX

PT    Enriched library fraction comprising molecules recovered by in vivo

PT    panning that selectively home to a selected organ or tissue useful for

PT    treating disease or in diagnostic methods -

XX

PS    Example 2; Column 17; 21pp; English.

XX

CC    The invention relates to an enriched library fraction containing  
CC    molecules that selectively home to a selected organ or tissue such as  
CC    brain, kidney or tumour recovered by in vivo panning. The invention  
CC    generally relates to the field of molecular medicine, drug delivery and  
CC    to a method of invivo panning for identifying a molecule that homes to a  
CC    specific organ. The molecules, e.g., peptides, peptidomimetics, proteins  
CC    and fragments of proteins contained in an enriched library fraction may  
CC    be administered to a subject as part of a pharmaceutical composition to  
CC    treat disease or in diagnostic methods. The present sequence is a  
CC    peptide from bacteriophage targetted to brain.

XX

SQ    Sequence    7 AA;

Query Match 100.0%; Score 43; DB 22; Length 7;  
Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
      |||||  
Db 1 CGVRLGC 7

RESULT 4

AAU10709

ID AAU10709 standard; peptide; 7 AA.

XX

AC AAU10709;

XX

DT 12-MAR-2002 (first entry)

XX

DE Brain homing peptide #6 useful for delivery of target molecules.

XX

KW Organ targeting; tissue targeting; cancer; tumour homing molecule;  
KW delivery of target molecule; brain homing peptide.

XX

OS Synthetic.

XX

PN US6306365-B1.

XX

PD 23-OCT-2001.

XX

PF 08-JAN-1999; 99US-0227906.

XX

PR 23-JUN-1997; 97US-0862855.

PR 11-SEP-1995; 95US-0526710.

PR 10-MAR-1997; 97US-0813273.

XX

PA (BURN-) BURNHAM INST.

XX

PI Ruoslahti E, Pasqualini R;

XX

DR WPI; 2002-040196/05.

XX

PT Recovering molecules that home to an organ or tissue, useful for  
PT identifying molecules that home to a specific organ or tissue, e.g.  
PT identifying a tumour homing molecule to identify the presence of cancer,  
PT by in vivo panning of a library -

XX

PS Example 2; Column 17; 21pp; English.

XX

CC The present invention relates to a method of recovering molecules that  
CC home to a selected organ or tissue. The method comprises administering  
CC to the subject the library of diverse molecules, collecting a sample of  
CC the selected organ or tissue (e.g. brain or kidney), and recovering from  
CC the sample several molecules that home to the selected organ or tissue.  
CC The method is useful for identifying molecules, particularly useful for  
CC screening large number of molecules (e.g. peptides), that home to a  
CC specific organ. The identified molecule is useful for e.g. raising an  
CC antibody specific for a target molecule, targeting a desired moiety

CC (e.g. drug, toxin or detectable label) to the selected organ.  
CC Specifically, the method is useful for identifying the presence of cancer  
CC in a subject by linking an appropriate moiety to a tumour homing  
CC molecule. The present method provides a direct means for identifying  
CC molecules that specifically home to a selected organ and, therefore  
CC provides a significant advantage over previous methods, which require  
CC that a molecule identified using an in vitro screening method  
CC subsequently be examined to determine if it maintains its specificity in  
CC vivo. AAU10704-AAU10723 represent brain homing peptides described in  
CC the present invention.

XX

SQ Sequence 7 AA;

Query Match 100.0%; Score 43; DB 23; Length 7;  
Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
|||  
Db 1 CGVRLGC 7

#### RESULT 5

AAO02039

ID AAO02039 standard; Protein; 61 AA.

XX

AC AAO02039;

XX

DT 06-NOV-2001 (first entry)

XX

DE Human polypeptide SEQ ID NO 15931.

XX

KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;

KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

KW tissue growth factor; immunomodulatory; cancer; leukaemia;

KW nervous system disorders; arthritis; inflammation.

XX

OS Homo sapiens.

XX

PN WO200164835-A2.

XX

PD 07-SEP-2001.

XX

PF 26-FEB-2001; 2001WO-US04927.

XX

PR 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Tang YT, Liu C, Drmanac RT;

XX

DR WPI; 2001-514838/56.

DR N-PSDB; AAI81970.

XX

PT Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune

PT disorders -  
 XX  
 PS Claim 20; SEQ ID NO 15931; 1399pp + Sequence Listing; English.  
 XX  
 CC The invention relates to human polynucleotides (AAI79941-AAI93841) and  
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 61 AA;

Query Match 83.7%; Score 36; DB 22; Length 61;  
 Best Local Similarity 85.7%; Pred. No. 53;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
 ||| |||  
 Db 53 CGVLLGC 59

# RESULT 6

AAG67507

ID AAG67507 standard; Protein; 159 AA.

XX

AC AAG67507;

XX

DT 26-NOV-2001 (first entry)

XX

DE Amino acid sequence of a human secreted polypeptide.

XX

KW Human; secreted polypeptide; nervous disease; muscular disease; tumour;

KW gastrointestinal ulceration; spinal cord disease; trachea disease;

KW thyroid gland disease; ovary disease; prostate disease; heart disease;

KW renal gland disease; small intestine disease; thymus disease;

KW lymph node disease; muscular system disease; colon disease;

KW lipase deficiency; cystic fibrosis; pancreatitis; clot formation;

KW myocardial infarction; angioplasty; liver disease; coagulation disorder;

KW microbial disease; immune disorder; inflammation; transplant rejection;

KW bone thickness; bone density; ferroxidase loss; apoptosis;

KW vascular smooth cell proliferation; vaccine.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Misc-difference 1

FT /note= "the nucleotides encoding this residue are  
 FT not given"



XX  
 PN WO200166690-A2.  
 XX  
 PD 13-SEP-2001.  
 XX  
 PF 05-MAR-2001; 2001WO-US07143.  
 XX  
 PR 06-MAR-2000; 2000US-0187107.  
 PR 13-MAR-2000; 2000US-0188916.  
 PR 03-OCT-2000; 2000US-0236874.  
 PR 03-OCT-2000; 2000US-0237846.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 XX  
 PI Agarwal P, Murdoch PR, Rizvi SK, Smith RF, Xiang Z;  
 XX  
 DR WPI; 2001-570768/64.  
 DR N-PSDB; AAH78199.  
 XX  
 PT Novel isolated secreted polypeptide useful for treating nervous and  
 PT muscular diseases, gastrointestinal ulceration, coagulation and immune  
 PT disorders, microbial diseases, inflammation and transplant rejection -  
 XX  
 PS Claim 1; Page 61; 102pp; English.  
 XX  
 CC The present sequence represents a human secreted polypeptide. The  
 CC secreted polypeptides and polynucleotides are useful for treating  
 CC nervous and muscular diseases, for inhibiting tumour formation and  
 CC metastasis, for treating gastrointestinal ulceration, for preventing  
 CC and treating diseases in spinal cord, thyroid gland, ovary, prostate,  
 CC renal gland, small intestine, heart, trachea, thymus, lymph node,  
 CC muscular system and colon, for treating lipase deficiency in cystic  
 CC fibrosis and pancreatitis, for treating undesirable clot formation  
 CC such as myocardial infarction, during angioplasty and all surgical  
 CC procedures that require decreased blood clot formation, for treating  
 CC liver diseases, coagulation disorders and microbial diseases, for  
 CC treating immune disorders, for treating inflammation and transplant  
 CC rejection, for enhancing bone thickness and increasing bone density,  
 CC for reducing the loss of essential ferroxidases, for suppressing  
 CC apoptosis, and for regulating vascular smooth cell proliferation. They  
 CC may also be used as vaccines.  
 XX  
 SQ Sequence 159 AA;  
  
 Query Match 83.7%; Score 36; DB 22; Length 159;  
 Best Local Similarity 71.4%; Pred. No. 1.2e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
 QY 1 CGVRLGC 7  
 || ||:  
 Db 132 CGCRMGC 138

# RESULT 7

AAB48195

ID AAB48195 standard; Protein; 434 AA.

XX  
 AC AAB48195;  
 XX  
 DT 02-APR-2001 (first entry)  
 XX  
 DE Drosophila mutant DIAP1 33-1S peptide.  
 XX  
 KW Drosophila inhibitor of apoptosis protein 1; DIAP1; DIAP1 6-3S; mutant;  
 KW DIAP1 45-2S; DIAP1 23-4S; DIAP1 11-3E; DIAP1 22-8S; DIAP1 21-4S; tumour;  
 KW DIAP1 33-1S; DIAP1 21-2S; DIAP1 41-8S; IAP.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 350  
 FT /note= "encoded by TAG"  
 FT Misc-difference 415  
 FT /note= "encoded by TGA"  
 FT Misc-difference 428  
 FT /note= "encoded by TGA"  
 XX  
 PN WO200075161-A2.  
 XX  
 PD 14-DEC-2000.  
 XX  
 PF 02-JUN-2000; 2000WO-US15278.  
 XX  
 PR 04-JUN-1999; 99US-0137624.  
 XX  
 PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
 XX  
 PI Steller H, McCall K, Goyal L, Agapite J;  
 XX  
 DR WPI; 2001-091199/10.  
 DR N-PSDB; AAC84527.  
 XX  
 PT New DNA composition for Drosophila inhibitor of apoptosis protein 1, is  
 PT useful for screening compounds that enhance or reduce apoptosis,  
 PT particularly for screening tumors that manifest mutations in homologs  
 PT to the apoptosis protein 1 gene -  
 XX  
 PS Disclosure; Fig 14; 49pp; English.  
 XX  
 CC The invention relates to novel mutant forms of Drosophila inhibitor of  
 CC apoptosis protein 1 (DIAP1). The mutants are DIAP1 6-3S, DIAP1 45-2S,  
 CC DIAP1 23-4S, DIAP1 11-3E, DIAP1 22-8S, DIAP1 21-4S, DIAP1 33-1S,  
 CC DIAP1 21-2S or DIAP1 41-8S and can be produced by standard recombinant  
 CC methodology. Compositions comprising the mutant sequences is useful in  
 CC screening assays, especially in a cell-free assay system for identifying  
 CC and testing DIAP1 pathway antagonists and agonists, screening for tumours  
 CC that manifest mutations in genes similar to, or homologous with, the  
 CC DIAP1 cDNA. It is also useful for screening compounds with agonistic or  
 CC antagonistic effects on apoptosis, particularly for compounds that exert  
 CC their effect at the level of IAPs (inhibitors of apoptosis protein). The  
 CC present sequence represents the mutant DIAP1 33-1S.  
 XX  
 SQ Sequence 434 AA;

Query Match 83.7%; Score 36; DB 22; Length 434;  
Best Local Similarity 71.4%; Pred. No. 2.8e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
||| :||  
Db 83 CGVEIGC 89

RESULT 8

ABB61858

ID ABB61858 standard; Protein; 438 AA.

XX

AC ABB61858;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster polypeptide SEQ ID NO 12366.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;  
pharmaceutical.

XX

OS Drosophila melanogaster.

XX

PN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US09231.

XX

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX

PA (PEKE ) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers EW;

XX

DR WPI; 2001-656860/75.

DR N-PSDB; ABL05961.

XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
genes from Drosophila and for elucidating cell signalling and cell-cell  
interactions -

XX

PS Disclosure; SEQ ID NO 12366; 21pp + Sequence Listing; English.

XX

CC The invention relates to an isolated nucleic acid detection reagent  
capable of detecting 1000 or more genes from Drosophila. The invention is  
useful in developmental biology and in elucidating cell signalling and  
cell-cell interactions in higher eukaryotes for the development of  
insecticides, therapeutics and pharmaceutical drugs. The invention  
discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
sequences (ABL01840-ABL16175) and the encoded proteins  
(ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed  
specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 438 AA;

Query Match 83.7%; Score 36; DB 22; Length 438;

Best Local Similarity 71.4%; Pred. No. 2.8e+02;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7

||| :||

Db 83 CGVEIGC 89

#### RESULT 9

ABB67347

ID ABB67347 standard; Protein; 438 AA.

XX

AC ABB67347;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster polypeptide SEQ ID NO 28833.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.

XX

OS Drosophila melanogaster.

XX

PN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US09231.

XX

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX

PA (PEKE ) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers EW;

XX

DR WPI; 2001-656860/75.

DR N-PSDB; ABL11450.

XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -

XX

PS Disclosure; SEQ ID NO 28833; 21pp + Sequence Listing; English.

XX

CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 438 AA;

Query Match 83.7%; Score 36; DB 22; Length 438;  
 Best Local Similarity 71.4%; Pred. No. 2.8e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
 ||| :||  
 Db 83 CGVEIGC 89

RESULT 10

AAB48188

ID AAB48188 standard; Protein; 438 AA.

XX

AC AAB48188;

XX

DT 02-APR-2001 (first entry)

XX

DE Drosophila wild-type DIAP1 peptide.

XX

KW Drosophila inhibitor of apoptosis protein 1; DIAP1; DIAP1 6-3S; mutant;

KW DIAP1 45-2S; DIAP1 23-4S; DIAP1 11-3E; DIAP1 22-8S; DIAP1 21-4S; tumour;

KW DIAP1 33-1S; DIAP1 21-2S; DIAP1 41-8S; IAP.

XX

OS Drosophila melanogaster.

XX

PN WO200075161-A2.

XX

PD 14-DEC-2000.

XX

PF 02-JUN-2000; 2000WO-US15278.

XX

PR 04-JUN-1999; 99US-0137624.

XX

PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.

XX

PI Steller H, McCall K, Goyal L, Agapite J;

XX

DR WPI; 2001-091199/10.

DR N-PSDB; AAC84520.

XX

PT New DNA composition for Drosophila inhibitor of apoptosis protein 1, is

PT useful for screening compounds that enhance or reduce apoptosis,

PT particularly for screening tumors that manifest mutations in homologs

PT to the apoptosis protein 1 gene -

XX

PS Disclosure; Fig 7; 49pp; English.

XX

CC The invention relates to novel mutant forms of Drosophila inhibitor of

CC apoptosis protein 1 (DIAP1). The mutants are DIAP1 6-3S, DIAP1 45-2S,

CC DIAP1 23-4S, DIAP1 11-3E, DIAP1 22-8S, DIAP1 21-4S, DIAP1 33-1S,

CC DIAP1 21-2S or DIAP1 41-8S and can be produced by standard recombinant  
 CC methodology. Compositions comprising the mutant sequences is useful in  
 CC screening assays, especially in a cell-free assay system for identifying  
 CC and testing DIAP1 pathway antagonists and agonists, screening for tumours  
 CC that manifest mutations in genes similar to, or homologous with, the  
 CC DIAP1 cDNA. It is also useful for screening compounds with agonistic or  
 CC antagonistic effects on apoptosis, particularly for compounds that exert  
 CC their effect at the level of IAPs (inhibitors of apoptosis protein). The  
 CC present sequence represents the wild-type DIAP1 peptide.  
 XX  
 SQ Sequence 438 AA;

Query Match 83.7%; Score 36; DB 22; Length 438;  
 Best Local Similarity 71.4%; Pred. No. 2.8e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
 ||| :||  
 Db 83 CGVEIGC 89

# RESULT 11

AAB48191

ID AAB48191 standard; Protein; 438 AA.

XX

AC AAB48191;

XX

DT 02-APR-2001 (first entry)

XX

DE Drosophila mutant DIAP1 23-4S peptide.

XX

KW Drosophila inhibitor of apoptosis protein 1; DIAP1; DIAP1 6-3S; mutant;  
 KW DIAP1 45-2S; DIAP1 23-4S; DIAP1 11-3E; DIAP1 22-8S; DIAP1 21-4S; tumour;  
 KW DIAP1 33-1S; DIAP1 21-2S; DIAP1 41-8S; IAP.

XX

OS Drosophila melanogaster.

XX

PN WO200075161-A2.

XX

PD 14-DEC-2000.

XX

PF 02-JUN-2000; 2000WO-US15278.

XX

PR 04-JUN-1999; 99US-0137624.

XX

PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.

XX

PI Steller H, McCall K, Goyal L, Agapite J;

XX

DR WPI; 2001-091199/10.

DR N-PSDB; AAC84523.

XX

PT New DNA composition for Drosophila inhibitor of apoptosis protein 1, is  
 PT useful for screening compounds that enhance or reduce apoptosis,  
 PT particularly for screening tumors that manifest mutations in homologs  
 PT to the apoptosis protein 1 gene -

XX

PS Disclosure; Fig 10; 49pp; English.

XX

CC The invention relates to novel mutant forms of Drosophila inhibitor of  
CC apoptosis protein 1 (DIAP1). The mutants are DIAP1 6-3S, DIAP1 45-2S,  
CC DIAP1 23-4S, DIAP1 11-3E, DIAP1 22-8S, DIAP1 21-4S, DIAP1 33-1S,  
CC DIAP1 21-2S or DIAP1 41-8S and can be produced by standard recombinant  
CC methodology. Compositions comprising the mutant sequences is useful in  
CC screening assays, especially in a cell-free assay system for identifying  
CC and testing DIAP1 pathway antagonists and agonists, screening for tumours  
CC that manifest mutations in genes similar to, or homologous with, the  
CC DIAP1 cDNA. It is also useful for screening compounds with agonistic or  
CC antagonistic effects on apoptosis, particularly for compounds that exert  
CC their effect at the level of IAPs (inhibitors of apoptosis protein). The  
CC present sequence represents the mutant DIAP1 23-4S.

XX

SQ Sequence 438 AA;

Query Match 83.7%; Score 36; DB 22; Length 438;

Best Local Similarity 71.4%; Pred. No. 2.8e+02;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7

||| :||

Db 83 CGVEIGC 89

## RESULT 12

AAB48192

ID AAB48192 standard; Protein; 438 AA.

XX

AC AAB48192;

XX

DT 02-APR-2001 (first entry)

XX

DE Drosophila mutant DIAP1 11-3E peptide.

XX

KW Drosophila inhibitor of apoptosis protein 1; DIAP1; DIAP1 6-3S; mutant;

KW DIAP1 45-2S; DIAP1 23-4S; DIAP1 11-3E; DIAP1 22-8S; DIAP1 21-4S; tumour;

KW DIAP1 33-1S; DIAP1 21-2S; DIAP1 41-8S; IAP.

XX

OS Drosophila melanogaster.

XX

PN WO200075161-A2.

XX

PD 14-DEC-2000.

XX

PF 02-JUN-2000; 2000WO-US15278.

XX

PR 04-JUN-1999; 99US-0137624.

XX

PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.

XX

PI Steller H, McCall K, Goyal L, Agapite J;

XX

DR WPI; 2001-091199/10.

DR N-PSDB; AAC84524.

XX

PT New DNA composition for Drosophila inhibitor of apoptosis protein 1, is  
PT useful for screening compounds that enhance or reduce apoptosis,  
PT particularly for screening tumors that manifest mutations in homologs  
PT to the apoptosis protein 1 gene -  
XX

PS Disclosure; Fig 11; 49pp; English.  
XX

CC The invention relates to novel mutant forms of Drosophila inhibitor of  
CC apoptosis protein 1 (DIAP1). The mutants are DIAP1 6-3S, DIAP1 45-2S,  
CC DIAP1 23-4S, DIAP1 11-3E, DIAP1 22-8S, DIAP1 21-4S, DIAP1 33-1S,  
CC DIAP1 21-2S or DIAP1 41-8S and can be produced by standard recombinant  
CC methodology. Compositions comprising the mutant sequences is useful in  
CC screening assays, especially in a cell-free assay system for identifying  
CC and testing DIAP1 pathway antagonists and agonists, screening for tumours  
CC that manifest mutations in genes similar to, or homologous with, the  
CC DIAP1 cDNA. It is also useful for screening compounds with agonistic or  
CC antagonistic effects on apoptosis, particularly for compounds that exert  
CC their effect at the level of IAPs (inhibitors of apoptosis protein). The  
CC present sequence represents the mutant DIAP1 11-3E.  
XX

SQ Sequence 438 AA;

Query Match 83.7%; Score 36; DB 22; Length 438;

Best Local Similarity 71.4%; Pred. No. 2.8e+02;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
||| :||  
Db 83 CGVEIGC 89

#### RESULT 13

AAB48193

ID AAB48193 standard; Protein; 438 AA.

XX

AC AAB48193;

XX

DT 02-APR-2001 (first entry)

XX

DE Drosophila mutant DIAP1 22-8S peptide.

XX

KW Drosophila inhibitor of apoptosis protein 1; DIAP1; DIAP1 6-3S; mutant;  
KW DIAP1 45-2S; DIAP1 23-4S; DIAP1 11-3E; DIAP1 22-8S; DIAP1 21-4S; tumour;  
KW DIAP1 33-1S; DIAP1 21-2S; DIAP1 41-8S; IAP.

XX

OS Drosophila melanogaster.

XX

PN WO200075161-A2.

XX

PD 14-DEC-2000.

XX

PF 02-JUN-2000; 2000WO-US15278.

XX

PR 04-JUN-1999; 99US-0137624.

XX

PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.

XX



PI Steller H, McCall K, Goyal L, Agapite J;

XX

DR WPI; 2001-091199/10.

DR N-PSDB; AAC84525.

XX

PT New DNA composition for Drosophila inhibitor of apoptosis protein 1, is  
PT useful for screening compounds that enhance or reduce apoptosis,  
PT particularly for screening tumors that manifest mutations in homologs  
PT to the apoptosis protein 1 gene -

XX

PS Disclosure; Fig 12; 49pp; English.

XX

CC The invention relates to novel mutant forms of Drosophila inhibitor of  
CC apoptosis protein 1 (DIAP1). The mutants are DIAP1 6-3S, DIAP1 45-2S,  
CC DIAP1 23-4S, DIAP1 11-3E, DIAP1 22-8S, DIAP1 21-4S, DIAP1 33-1S,  
CC DIAP1 21-2S or DIAP1 41-8S and can be produced by standard recombinant  
CC methodology. Compositions comprising the mutant sequences is useful in  
CC screening assays, especially in a cell-free assay system for identifying  
CC and testing DIAP1 pathway antagonists and agonists, screening for tumours  
CC that manifest mutations in genes similar to, or homologous with, the  
CC DIAP1 cDNA. It is also useful for screening compounds with agonistic or  
CC antagonistic effects on apoptosis, particularly for compounds that exert  
CC their effect at the level of IAPs (inhibitors of apoptosis protein). The  
CC present sequence represents the mutant DIAP1 22-8S.

XX

SQ Sequence 438 AA;

Query Match 83.7%; Score 36; DB 22; Length 438;

Best Local Similarity 71.4%; Pred. No. 2.8e+02;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7

||| :||

Db 83 CGVEIGC 89

#### RESULT 14

AAB48194

ID AAB48194 standard; Protein; 438 AA.

XX

AC AAB48194;

XX

DT 02-APR-2001 (first entry)

XX

DE Drosophila mutant DIAP1 21-4S peptide.

XX

KW Drosophila inhibitor of apoptosis protein 1; DIAP1; DIAP1 6-3S; mutant;  
KW DIAP1 45-2S; DIAP1 23-4S; DIAP1 11-3E; DIAP1 22-8S; DIAP1 21-4S; tumour;  
KW DIAP1 33-1S; DIAP1 21-2S; DIAP1 41-8S; IAP.

XX

OS Drosophila melanogaster.

XX

PN WO200075161-A2.

XX

PD 14-DEC-2000.

XX

PF 02-JUN-2000; 2000WO-US15278.

XX  
 PR 04-JUN-1999; 99US-0137624.  
 XX  
 PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
 XX  
 PI Steller H, McCall K, Goyal L, Agapite J;  
 XX  
 DR WPI; 2001-091199/10.  
 DR N-PSDB; AAC84526.  
 XX  
 PT New DNA composition for Drosophila inhibitor of apoptosis protein 1, is  
 PT useful for screening compounds that enhance or reduce apoptosis,  
 PT particularly for screening tumors that manifest mutations in homologs  
 PT to the apoptosis protein 1 gene -  
 XX  
 PS Disclosure; Fig 13; 49pp; English.  
 XX  
 CC The invention relates to novel mutant forms of Drosophila inhibitor of  
 CC apoptosis protein 1 (DIAP1). The mutants are DIAP1 6-3S, DIAP1 45-2S,  
 CC DIAP1 23-4S, DIAP1 11-3E, DIAP1 22-8S, DIAP1 21-4S, DIAP1 33-1S,  
 CC DIAP1 21-2S or DIAP1 41-8S and can be produced by standard recombinant  
 CC methodology. Compositions comprising the mutant sequences is useful in  
 CC screening assays, especially in a cell-free assay system for identifying  
 CC and testing DIAP1 pathway antagonists and agonists, screening for tumours  
 CC that manifest mutations in genes similar to, or homologous with, the  
 CC DIAP1 cDNA. It is also useful for screening compounds with agonistic or  
 CC antagonistic effects on apoptosis, particularly for compounds that exert  
 CC their effect at the level of IAPs (inhibitors of apoptosis protein). The  
 CC present sequence represents the mutant DIAP1 21-4S.  
 XX  
 SQ Sequence 438 AA;

Query Match 83.7%; Score 36; DB 22; Length 438;  
 Best Local Similarity 71.4%; Pred. No. 2.8e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
 ||| :||  
 Db 83 CGVEIGC 89

# RESULT 15 AAB48196

ID AAB48196 standard; Protein; 438 AA.  
 XX  
 AC AAB48196;  
 XX  
 DT 02-APR-2001 (first entry)  
 XX  
 DE Drosophila mutant DIAP1 21-2S peptide.  
 XX  
 KW Drosophila inhibitor of apoptosis protein 1; DIAP1; DIAP1 6-3S; mutant;  
 KW DIAP1 45-2S; DIAP1 23-4S; DIAP1 11-3E; DIAP1 22-8S; DIAP1 21-4S; tumour;  
 KW DIAP1 33-1S; DIAP1 21-2S; DIAP1 41-8S; IAP.  
 XX  
 OS Drosophila melanogaster.  
 XX

PN WO200075161-A2.  
 XX  
 PD 14-DEC-2000.  
 XX  
 PF 02-JUN-2000; 2000WO-US15278.  
 XX  
 PR 04-JUN-1999; 99US-0137624.  
 XX  
 PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
 XX  
 PI Steller H, McCall K, Goyal L, Agapite J;  
 XX  
 DR WPI; 2001-091199/10.  
 DR N-PSDB; AAC84528.  
 XX  
 PT New DNA composition for Drosophila inhibitor of apoptosis protein 1, is  
 PT useful for screening compounds that enhance or reduce apoptosis,  
 PT particularly for screening tumors that manifest mutations in homologs  
 PT to the apoptosis protein 1 gene -  
 XX  
 PS Disclosure; Fig 15; 49pp; English.  
 XX  
 CC The invention relates to novel mutant forms of Drosophila inhibitor of  
 CC apoptosis protein 1 (DIAP1). The mutants are DIAP1 6-3S, DIAP1 45-2S,  
 CC DIAP1 23-4S, DIAP1 11-3E, DIAP1 22-8S, DIAP1 21-4S, DIAP1 33-1S,  
 CC DIAP1 21-2S or DIAP1 41-8S and can be produced by standard recombinant  
 CC methodology. Compositions comprising the mutant sequences is useful in  
 CC screening assays, especially in a cell-free assay system for identifying  
 CC and testing DIAP1 pathway antagonists and agonists, screening for tumours  
 CC that manifest mutations in genes similar to, or homologous with, the  
 CC DIAP1 cDNA. It is also useful for screening compounds with agonistic or  
 CC antagonistic effects on apoptosis, particularly for compounds that exert  
 CC their effect at the level of IAPs (inhibitors of apoptosis protein). The  
 CC present sequence represents the mutant DIAP1 21-2S.  
 XX  
 SQ Sequence 438 AA;

Query Match 83.7%; Score 36; DB 22; Length 438;  
 Best Local Similarity 71.4%; Pred. No. 2.8e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
 ||| :||  
 Db 83 CGVEIGC 89

Search completed: November 13, 2003, 09:45:26  
 Job time : 24.5521 secs

GenCore version 5.1.6  
 Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:45:35 ; Search time 14.5104 Seconds  
 (without alignments)

88.069 Million cell updates/sec

Title: US-09-228-866-6  
Perfect score: 43  
Sequence: 1 CGVRLGC 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description	
1	38	88.4	67	12	US-10-029-386-29430	Sequence 29430, A
2	36	83.7	109	10	US-09-965-967-19	Sequence 19, Appl
3	36	83.7	159	12	US-10-221-097-29	Sequence 29, Appl
4	36	83.7	172	15	US-10-041-859-13	Sequence 13, Appl
5	36	83.7	674	15	US-10-173-123-9	Sequence 9, Appli
6	36	83.7	681	15	US-10-173-123-7	Sequence 7, Appli
7	36	83.7	738	15	US-10-173-123-13	Sequence 13, Appl
8	36	83.7	745	15	US-10-173-123-11	Sequence 11, Appl

9	35	81.4	21	9	US-09-864-761-41531	Sequence 41531, A
10	35	81.4	232	10	US-09-738-626-4827	Sequence 4827, Ap
11	34	79.1	219	11	US-09-975-719-415	Sequence 415, App
12	33	76.7	35	10	US-09-843-676-178	Sequence 178, App
13	33	76.7	35	11	US-09-438-486-178	Sequence 178, App
14	33	76.7	35	15	US-10-053-758-178	Sequence 178, App
15	33	76.7	35	15	US-10-054-295-178	Sequence 178, App
16	33	76.7	35	15	US-10-054-611-178	Sequence 178, App
17	33	76.7	664	15	US-10-119-988-12	Sequence 12, Appl
18	33	76.7	8360	12	US-10-132-134-34	Sequence 34, Appl
19	32	74.4	44	15	US-10-083-357-822	Sequence 822, App
20	32	74.4	71	15	US-10-083-357-763	Sequence 763, App
21	32	74.4	146	10	US-09-738-626-5431	Sequence 5431, Ap
22	32	74.4	194	12	US-10-216-163-32	Sequence 32, Appl
23	32	74.4	194	12	US-10-218-765-32	Sequence 32, Appl
24	32	74.4	194	12	US-10-219-063-32	Sequence 32, Appl
25	32	74.4	194	12	US-10-219-066-32	Sequence 32, Appl
26	32	74.4	194	12	US-10-219-067-32	Sequence 32, Appl
27	32	74.4	194	12	US-10-219-068-32	Sequence 32, Appl
28	32	74.4	194	12	US-10-219-069-32	Sequence 32, Appl
29	32	74.4	194	12	US-10-219-073-32	Sequence 32, Appl
30	32	74.4	194	12	US-10-219-475-32	Sequence 32, Appl
31	32	74.4	194	12	US-10-219-480-32	Sequence 32, Appl
32	32	74.4	194	12	US-10-219-483-32	Sequence 32, Appl
33	32	74.4	194	12	US-10-219-525-32	Sequence 32, Appl
34	32	74.4	194	12	US-10-219-526-32	Sequence 32, Appl
35	32	74.4	194	12	US-10-219-530-32	Sequence 32, Appl
36	32	74.4	194	12	US-10-219-531-32	Sequence 32, Appl
37	32	74.4	194	12	US-10-219-532-32	Sequence 32, Appl
38	32	74.4	194	12	US-10-219-533-32	Sequence 32, Appl
39	32	74.4	194	12	US-10-230-437-32	Sequence 32, Appl
40	32	74.4	194	12	US-10-232-228-32	Sequence 32, Appl
41	32	74.4	194	15	US-10-227-884-32	Sequence 32, Appl
42	32	74.4	194	15	US-10-230-163-32	Sequence 32, Appl
43	32	74.4	194	15	US-10-230-338-32	Sequence 32, Appl
44	32	74.4	194	15	US-10-218-631-32	Sequence 32, Appl
45	32	74.4	194	15	US-10-230-414-32	Sequence 32, Appl

#### ALIGNMENTS

##### RESULT 1

US-10-029-386-29430

; Sequence 29430, Application US/10029386

; Publication No. US20030194704A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES  
USEFUL FOR GENE

; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO

; FILE REFERENCE: AEOMICA-X-2

; CURRENT APPLICATION NUMBER: US/10/029,386

; CURRENT FILING DATE: 2001-12-20

; NUMBER OF SEQ ID NOS: 34288

```
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 29430
; LENGTH: 67
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR5.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.87
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1
; OTHER INFORMATION: SWISSPROT HIT: P57074, EVALUE 2.40e+00
US-10-029-386-29430
```

```
Query Match          88.4%;  Score 38;  DB 12;  Length 67;
Best Local Similarity 71.4%;  Pred. No. 10;
Matches      5;  Conservative    2;  Mismatches    0;  Indels      0;  Gaps      0;
```

```
Qy      1 CGVRLGC 7
        ||::|||
Db      32 CGIQLGC 38
```

#### RESULT 2

```
US-09-965-967-19
; Sequence 19, Application US/09965967
; Patent No. US20020177557A1
; GENERAL INFORMATION:
; APPLICANT: Shi, Yigong
; TITLE OF INVENTION: Compositions And Methods For Regulating Apoptosis
; FILE REFERENCE: PU-0031 (01-1739-1)
; CURRENT APPLICATION NUMBER: US/09/965,967
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: 60/236,574
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/256,830
; PRIOR FILING DATE: 2000-12-20
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-965-967-19
```

```
Query Match          83.7%;  Score 36;  DB 10;  Length 109;
Best Local Similarity 71.4%;  Pred. No. 35;
Matches      5;  Conservative    1;  Mismatches    1;  Indels      0;  Gaps      0;
```

```
Qy      1 CGVRLGC 7
        ||| :||
Db      54 CGVEIGC 60
```

#### RESULT 3

```
US-10-221-097-29
```

```
; Sequence 29, Application US/10221097
; Publication No. US20030144476A1
; GENERAL INFORMATION:
; APPLICANT: Agarwal, Pankaj
; APPLICANT: Murdock, Paul R.
; APPLICANT: Rizvi, Safia K.
; APPLICANT: Smith, Randall F.
; APPLICANT: Xiang, Zhaoying
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50016
; CURRENT APPLICATION NUMBER: US/10/221,097
; CURRENT FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: PCT/US01/07143
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/187,107
; PRIOR FILING DATE: 2000-03-06
; PRIOR APPLICATION NUMBER: 60/236,874
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/188,916
; PRIOR FILING DATE: 2000-03-13
; PRIOR APPLICATION NUMBER: 60/237,846
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 29
;   LENGTH: 159
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-221-097-29
```

```
Query Match          83.7%;  Score 36;  DB 12;  Length 159;
Best Local Similarity 71.4%;  Pred. No. 50;
Matches      5;  Conservative    1;  Mismatches    1;  Indels      0;  Gaps      0;
```

```
Qy      1 CGVRLGC 7
        || |:||
Db      132 CGCRMGC 138
```

```
RESULT 4
US-10-041-859-13
; Sequence 13, Application US/10041859
; Publication No. US20030049796A1
; GENERAL INFORMATION:
; APPLICANT: HUANG, QIHONG
; APPLICANT: REED, JOHN C.
; APPLICANT: DEVERAUX, QUINN L.
; APPLICANT: MAEDA, SUSUMU
; TITLE OF INVENTION: INHIBITOR OF APOPTOSIS PROTEINS AND NUCLEIC ACIDS AND
; TITLE OF INVENTION: METHODS FOR MAKING AND USING THEM
; FILE REFERENCE: 087102/027 2537
; CURRENT APPLICATION NUMBER: US/10/041,859
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: 60/260,478
; PRIOR FILING DATE: 2001-01-08
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.1
```

; SEQ ID NO 13  
; LENGTH: 172  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-10-041-859-13

Query Match 83.7%; Score 36; DB 15; Length 172;  
Best Local Similarity 71.4%; Pred. No. 54;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
||| :||  
Db 40 CGVEIGC 46

RESULT 5

US-10-173-123-9  
; Sequence 9, Application US/10173123  
; Publication No. US20030027301A1  
; GENERAL INFORMATION:  
; APPLICANT: Hu, Yi  
; APPLICANT: Nepomnichy, Boris  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Mathur, Brian  
; APPLICANT: Friddle, Carl Johan  
; TITLE OF INVENTION: No. US20030027301A1el Human Transporter Proteins and  
; TITLE OF INVENTION: Polynucleotides Encoding the Same  
; FILE REFERENCE: LEX-0358-USA  
; CURRENT APPLICATION NUMBER: US/10/173,123  
; CURRENT FILING DATE: 2002-06-14  
; PRIOR APPLICATION NUMBER: US 60/298,241  
; PRIOR FILING DATE: 2001-06-14  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 9  
; LENGTH: 674  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-173-123-9

Query Match 83.7%; Score 36; DB 15; Length 674;  
Best Local Similarity 71.4%; Pred. No. 1.9e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
|| |:||  
Db 352 CGARVGC 358

RESULT 6

US-10-173-123-7  
; Sequence 7, Application US/10173123  
; Publication No. US20030027301A1  
; GENERAL INFORMATION:  
; APPLICANT: Hu, Yi  
; APPLICANT: Nepomnichy, Boris  
; APPLICANT: Turner, C. Alexander Jr.



```
; APPLICANT: Mathur, Brian
; APPLICANT: Friddle, Carl Johan
; TITLE OF INVENTION: No. US20030027301A1el Human Transporter Proteins and
; TITLE OF INVENTION: Polynucleotides Encoding the Same
; FILE REFERENCE: LEX-0358-USA
; CURRENT APPLICATION NUMBER: US/10/173,123
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: US 60/298,241
; PRIOR FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 681
; TYPE: PRT
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 124, 152
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-173-123-7
```

```
Query Match          83.7%; Score 36; DB 15; Length 681;
Best Local Similarity 71.4%; Pred. No. 2e+02;
Matches      5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy      1 CGVRLGC 7
        || |:||
Db      359 CGARVGC 365
```

#### RESULT 7

```
US-10-173-123-13
; Sequence 13, Application US/10173123
; Publication No. US20030027301A1
; GENERAL INFORMATION:
; APPLICANT: Hu, Yi
; APPLICANT: Nepomnichy, Boris
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Mathur, Brian
; APPLICANT: Friddle, Carl Johan
; TITLE OF INVENTION: No. US20030027301A1el Human Transporter Proteins and
; TITLE OF INVENTION: Polynucleotides Encoding the Same
; FILE REFERENCE: LEX-0358-USA
; CURRENT APPLICATION NUMBER: US/10/173,123
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: US 60/298,241
; PRIOR FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 738
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-173-123-13
```

```
Query Match          83.7%; Score 36; DB 15; Length 738;
Best Local Similarity 71.4%; Pred. No. 2.1e+02;
```

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
|| |:||  
Db 416 CGARVGC 422

RESULT 8

US-10-173-123-11  
; Sequence 11, Application US/10173123  
; Publication No. US20030027301A1  
; GENERAL INFORMATION:  
; APPLICANT: Hu, Yi  
; APPLICANT: Nepomnichy, Boris  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Mathur, Brian  
; APPLICANT: Friddle, Carl Johan  
; TITLE OF INVENTION: No. US20030027301A1el Human Transporter Proteins and  
; TITLE OF INVENTION: Polynucleotides Encoding the Same  
; FILE REFERENCE: LEX-0358-USA  
; CURRENT APPLICATION NUMBER: US/10/173,123  
; CURRENT FILING DATE: 2002-06-14  
; PRIOR APPLICATION NUMBER: US 60/298,241  
; PRIOR FILING DATE: 2001-06-14  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 11  
; LENGTH: 745  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-173-123-11

Query Match 83.7%; Score 36; DB 15; Length 745;  
Best Local Similarity 71.4%; Pred. No. 2.1e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
|| |:||  
Db 423 CGARVGC 429

RESULT 9

US-09-864-761-41531  
; Sequence 41531, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES  
USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aeomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312

```

; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 41531
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC016057.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.3
US-09-864-761-41531

```

```

Query Match          81.4%;  Score 35;  DB 9;  Length 21;
Best Local Similarity 71.4%;  Pred. No. 11;
Matches      5;  Conservative 1;  Mismatches 1;  Indels 0;  Gaps 0;

```

Qy 1 CGVRLGC 7  
 ||: |||  
Db 2 CGILLGC 8

RESULT 10

US-09-738-626-4827  
; Sequence 4827, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAOKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 4827  
; LENGTH: 232  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-4827

Query Match 81.4%; Score 35; DB 10; Length 232;  
Best Local Similarity 85.7%; Pred. No. 1.1e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
 |||| ||  
Db 50 CGVRDGC 56

RESULT 11

US-09-975-719-415  
; Sequence 415, Application US/09975719  
; Publication No. US20030022349A1  
; GENERAL INFORMATION:  
; APPLICANT: Ausubel, Frederick M.  
; APPLICANT: Rahme, Laurence G.  
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID  
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF  
; FILE REFERENCE: 00786/361003

; CURRENT APPLICATION NUMBER: US/09/975,719  
; CURRENT FILING DATE: 2001-10-10  
; PRIOR APPLICATION NUMBER: US 09/199,637  
; PRIOR FILING DATE: 1998-11-25  
; PRIOR APPLICATION NUMBER: US 60/066,517  
; PRIOR FILING DATE: 1997-11-25  
; NUMBER OF SEQ ID NOS: 437  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 415  
; LENGTH: 219  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-975-719-415

Query Match 79.1%; Score 34; DB 11; Length 219;  
Best Local Similarity 85.7%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
| | | | |  
Db 63 CGVRLCC 69

RESULT 12

US-09-843-676-178

; Sequence 178, Application US/09843676  
; Patent No. US20020164786A1

; GENERAL INFORMATION:

; APPLICANT: Cech, Thomas R.  
; Lingner, Joachim  
; Nakamura, Toru  
; Chapman, Karen B.  
; Morin, Gregg B.  
; Harley, Calvin  
; Andrews, William H.

; TITLE OF INVENTION: No. US20020164786A1e1 Telomerase

; NUMBER OF SEQUENCES: 225

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, 8th Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94111

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/843,676  
; FILING DATE: 26-Apr-2001  
; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/854,050  
; FILING DATE: 09-MAY-1997  
; APPLICATION NUMBER: US 08/846,017

```

; FILING DATE: 25-APR-1997
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015389-002930US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 178:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 178:

```

US-09-843-676-178

```

Query Match          76.7%;  Score 33;  DB 10;  Length 35;
Best Local Similarity 71.4%;  Pred. No. 41;
Matches      5;  Conservative    0;  Mismatches    2;  Indels      0;  Gaps      0;

```

```

Qy      1 CGVRLGC 7
        ||  |||
Db      2 CGTALGC 8

```

# RESULT 13

US-09-438-486-178

```

; Sequence 178, Application US/09438486
; Publication No. US20030009019A1
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Harley, Calvin
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: No. US20030009019A1el Telomerase
; NUMBER OF SEQUENCES: 223
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/438,486
; FILING DATE: 12-NOV-1999
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/851,843
; FILING DATE: 06-MAY-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015389-002931US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 178:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-438-486-178

```

```

Query Match          76.7%; Score 33; DB 11; Length 35;
Best Local Similarity 71.4%; Pred. No. 41;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy      1 CGVRLGC 7
        ||  |||
Db      2 CGTALGC 8

```

```

RESULT 14
US-10-053-758-178
; Sequence 178, Application US/10053758
; Publication No. US20030032075A1
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
;           Lingner, Joachim
;           Nakamura, Toru
;           Chapman, Karen B.
;           Morin, Gregg B.
;           Harley, Calvin
;           Andrews, William H.

```

```

; TITLE OF INVENTION: No. US20030032075A1el Telomerase
; NUMBER OF SEQUENCES: 225
; CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Townsend and Townsend and Crew LLP
;     STREET: Two Embarcadero Center, 8th Floor
;     CITY: San Francisco
;     STATE: California
;     COUNTRY: United States of America
;     ZIP: 94111
; COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/10/053,758
;     FILING DATE: 18-Jan-2002
;     CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US/08/854,050
;     FILING DATE: 09-MAY-1997
;     APPLICATION NUMBER: US 08/851,843
;     FILING DATE: 06-MAY-1997
;     APPLICATION NUMBER: US 08/846,017
;     FILING DATE: 25-APR-1997
;     APPLICATION NUMBER: US 08/844,419
;     FILING DATE: 18-APR-1997
;     APPLICATION NUMBER: US 08/724,643
;     FILING DATE: 01-OCT-1996
; ATTORNEY/AGENT INFORMATION:
;     NAME: Apple, Randolph T.
;     REGISTRATION NUMBER: 36,429
;     REFERENCE/DOCKET NUMBER: 015389-002930US
; TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (415) 576-0200
;     TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 178:
;     SEQUENCE CHARACTERISTICS:
;         LENGTH: 35 amino acids
;         TYPE: amino acid
;         STRANDEDNESS: <Unknown>
;         TOPOLOGY: linear
;     MOLECULE TYPE: peptide
;     SEQUENCE DESCRIPTION: SEQ ID NO: 178:
US-10-053-758-178

```

```

Query Match          76.7%;  Score 33;  DB 15;  Length 35;
Best Local Similarity 71.4%;  Pred. No. 41;
Matches      5;  Conservative      0;  Mismatches      2;  Indels      0;  Gaps      0;

```

```

Qy      1 CGVRLGC 7
        ||  |||
Db      2 CGTALGC 8

```

```

RESULT 15
US-10-054-295-178

```



```

; Sequence 178, Application US/10054295
; Publication No. US20030044953A1
; GENERAL INFORMATION:
;   APPLICANT: Cech, Thomas R.
;               Lingner, Joachim
;               Nakamura, Toru
;               Chapman, Karen B.
;               Morin, Gregg B.
;               Harley, Calvin
;               Andrews, William H.
;   TITLE OF INVENTION: No. US20030044953A1el Telomerase
;   NUMBER OF SEQUENCES: 225
;   CORRESPONDENCE ADDRESS:
;       ADDRESSEE: Townsend and Townsend and Crew LLP
;       STREET: Two Embarcadero Center, 8th Floor
;       CITY: San Francisco
;       STATE: California
;       COUNTRY: United States of America
;       ZIP: 94111
;   COMPUTER READABLE FORM:
;       MEDIUM TYPE: Floppy disk
;       COMPUTER: IBM PC compatible
;       OPERATING SYSTEM: PC-DOS/MS-DOS
;       SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/10/054,295
;       FILING DATE: 18-Jan-2002
;       CLASSIFICATION: 536
;   PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 08/854,050
;       FILING DATE: <Unknown>
;       APPLICATION NUMBER: US 08/846,017
;       FILING DATE: 25-APR-1997
;       APPLICATION NUMBER: US 08/844,419
;       FILING DATE: 18-APR-1997
;       APPLICATION NUMBER: US 08/724,643
;       FILING DATE: 01-OCT-1996
;   ATTORNEY/AGENT INFORMATION:
;       NAME: Apple, Randolph T.
;       REGISTRATION NUMBER: 36,429
;       REFERENCE/DOCKET NUMBER: 015389-002930US
;   TELECOMMUNICATION INFORMATION:
;       TELEPHONE: (415) 576-0200
;       TELEFAX: (415) 576-0300
;   INFORMATION FOR SEQ ID NO: 178:
;       SEQUENCE CHARACTERISTICS:
;           LENGTH: 35 amino acids
;           TYPE: amino acid
;           STRANDEDNESS: <Unknown>
;           TOPOLOGY: linear
;       MOLECULE TYPE: peptide
;       SEQUENCE DESCRIPTION: SEQ ID NO: 178:
US-10-054-295-178

```

```

Query Match          76.7%;  Score 33;  DB 15;  Length 35;
Best Local Similarity 71.4%;  Pred. No. 41;
Matches      5;  Conservative    0;  Mismatches    2;  Indels      0;  Gaps      0;

```

Qy	1	CGVRLGC	7
Db	2	CGTALGC	8

Search completed: November 13, 2003, 09:58:28  
Job time : 15.5104 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:38:30 ; Search time 7.29167 Seconds  
(without alignments)  
92.322 Million cell updates/sec

```
Title:      US-09-228-866-6
Perfect score: 43
Sequence:   1 CGVRLGC 7
```

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```
Database :      PIR_76:*
1:  pir1:*
2:  pir2:*
3:  pir3:*
4:  pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result	Query					Description
No.	Score	Match	Length	DB	ID	
1	36	83.7	397	2	D72220	probable aspartate
2	36	83.7	662	2	A37226	glucose transport
3	36	83.7	1331	1	XORTDH	xanthine dehydroge
4	36	83.7	1335	1	XOMSDH	xanthine dehydroge
5	36	83.7	1335	2	S07245	xanthine dehydroge

6	36	83.7	1342	2	A31946	xanthine dehydroge
7	36	83.7	1353	2	JQ0407	xanthine dehydroge
8	36	83.7	1358	1	XOCHDH	xanthine dehydroge
9	34	79.1	132	2	A30978	MSEL neurophysin-c
10	34	79.1	161	2	S14480	arginine-vasotocin
11	34	79.1	228	2	B83583	dethiobiotin synth
12	34	79.1	267	2	B82694	copper homeostasis
13	34	79.1	382	2	T01943	hypothetical prote
14	34	79.1	758	2	S37855	hypothetical prote
15	34	79.1	772	2	T02805	chloride channel p
16	34	79.1	1402	2	I46707	translation initia
17	33	76.7	152	2	A32669	vasotocin 1 / neur
18	33	76.7	184	2	I39920	probable O-acetyl
19	33	76.7	262	2	T31253	1,6-dihydroxycyclo
20	33	76.7	605	2	A36361	glucose transport
21	33	76.7	664	2	S59638	glucose transport
22	33	76.7	664	2	A33545	Na+/glucose cotran
23	33	76.7	664	2	S59637	glucose transport
24	33	76.7	837	2	D84428	hypothetical prote
25	33	76.7	1565	2	AD2135	polyketide synthas
26	32	74.4	33	2	D61563	neurophysin 1 - ch
27	32	74.4	93	2	JS0301	neurophysin 1 - os
28	32	74.4	93	2	A60357	neurophysin 1 - go
29	32	74.4	122	2	S62781	phospholipase A2 (
30	32	74.4	125	2	A29879	mesotocin / neurop
31	32	74.4	311	2	AH3154	conserved hypothet
32	32	74.4	326	2	A99133	hypothetical prote
33	32	74.4	327	2	C72089	ABC transporter, p
34	32	74.4	327	2	C86534	solute protein bin
35	32	74.4	532	2	T07903	tubulin delta chai
36	32	74.4	605	2	E71253	alanine-tRNA ligas
37	32	74.4	694	2	D86615	polyribonucleotide
38	32	74.4	694	2	G72009	polyribonucleotide
39	32	74.4	718	2	A56851	Na+/myo-inositol c
40	32	74.4	718	2	A42163	Na+/myo-inositol c
41	32	74.4	719	2	G81332	polyribonucleotide
42	32	74.4	840	2	D82615	hypothetical prote
43	31.5	73.3	74	2	E87180	hypothetical prote
44	31	72.1	92	1	NFHO2	neurophysin 2 - ho
45	31	72.1	105	1	NFHO1	oxytocin / neuroph

#### ALIGNMENTS

##### RESULT 1

D72220

probable aspartate transaminase (EC 2.6.1.1) TM1698 [similarity] - *Thermotoga maritima* (strain MSB8)

C;Species: *Thermotoga maritima*

C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000

C;Accession: D72220

R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, E.K.; Peterson, J.D.; Nelson, W.C.; Ketchum, K.A.; McDonald, L.; Utterback, T.R.; Malek, J.A.; Linher, K.D.; Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; Heidelberg, J.;

Sutton, G.G.; Fleischmann, R.D.; White, O.; Salzberg, S.L.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Nature 399, 323-329, 1999

A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*.

A;Reference number: A72200; MUID:99287316; PMID:10360571

A;Accession: D72220

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-397 <ARN>

A;Cross-references: GB:AE001810; GB:AE000512; NID:g4982271; PIDN:AAD36765.1;

PID:g4982275; TIGR:TM1698

A;Experimental source: strain MSB8

C;Genetics:

A;Gene: TM1698

C;Superfamily: aspartate transaminase

C;Keywords: aminotransferase

Query Match 83.7%; Score 36; DB 2; Length 397;

Best Local Similarity 71.4%; Pred. No. 34;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7

|| |:||

Db 237 CGARVGC 243

## RESULT 2

A37226

glucose transport protein - rabbit

N;Alternate names: sodium/D-glucose cotransporter

C;Species: *Oryctolagus cuniculus* (domestic rabbit)

C;Date: 30-Dec-1991 #sequence\_revision 01-Mar-1996 #text\_change 20-Aug-1999

C;Accession: S00515; S15974; A37226

R;Hediger, M.A.; Coady, M.J.; Ikeda, T.S.; Wright, E.M.

Nature 330, 379-381, 1987

A;Title: Expression cloning and cDNA sequencing of the Na/glucose co-transporter.

A;Reference number: S00515; MUID:88065856; PMID:2446136

A;Accession: S00515

A;Molecule type: mRNA

A;Residues: 1-662 <HED>

A;Cross-references: EMBL:X06419; NID:g1640; PIDN:CAA29727.1; PID:g1641

R;Morrison, A.I.; Panayotova-Heiermann, M.; Feigl, G.; Schoelermann, B.; Kinne, R.K.H.

Biochim. Biophys. Acta 1089, 121-123, 1991

A;Title: Sequence comparison of the sodium-D-glucose cotransport systems in rabbit renal and intestinal epithelia.

A;Reference number: S15974; MUID:91223090; PMID:2025641

A;Accession: S15974

A;Molecule type: mRNA

A;Residues: 1-662 <MOR>

A;Cross-references: EMBL:X55355; NID:g1716; PIDN:CAA39040.1; PID:g1717

R;Coady, M.J.; Pajor, A.M.; Wright, E.M.

Am. J. Physiol. 259, C605-C610, 1990

A;Title: Sequence homologies among intestinal and renal Na(+)/glucose cotransporters.

A;Reference number: A37226; MUID:91023017; PMID:2221040  
A;Accession: A37226  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 178-662 <COA>  
A;Cross-references: GB:X06419  
A;Experimental source: renal cortex  
C;Superfamily: proline carrier protein

Query Match 83.7%; Score 36; DB 2; Length 662;  
Best Local Similarity 71.4%; Pred. No. 50;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
|||:|||  
Db 355 CGTRVGC 361

### RESULT 3

#### XORTDH

xanthine dehydrogenase (EC 1.1.1.204) / xanthine oxidase (EC 1.1.3.22) - rat  
N;Alternate names: hypoxanthine oxidase

C;Species: Rattus norvegicus (Norway rat)

C;Date: 30-Apr-1991 #sequence\_revision 07-Feb-1997 #text\_change 19-Jan-2001

C;Accession: A37810; S45259; S45260; S71397; I58308

R;Amaya, Y.; Yamazaki, K.; Sato, M.; Noda, K.; Nishino, T.; Nishino, T.  
J. Biol. Chem. 265, 14170-14175, 1990

A;Title: Proteolytic conversion of xanthine dehydrogenase from the NAD-dependent type to the O-2-dependent type. Amino acid sequence of rat liver xanthine dehydrogenase and identification of the cleavage sites of the enzyme protein during irreversible conversion by trypsin.

A;Reference number: A37810; MUID:90354396; PMID:2387845

A;Accession: A37810

A;Molecule type: mRNA

A;Residues: 1-478,491-493,'Q',495-1331 <AMA>

A;Cross-references: GB:J05579; NID:g207686

A;Note: parts of this sequence, including the amino end of the mature protein, were determined by protein sequencing

R;Chow, C.W.; Clark, M.; Rinaldo, J.; Chalkley, R.  
Nucleic Acids Res. 22, 1846-1854, 1994

A;Title: Identification of the rat xanthine dehydrogenase/oxidase promoter.

A;Reference number: I58308; MUID:94268906; PMID:8208609

A;Accession: S45259

A;Status: translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 476-494 <RES>

A;Cross-references: EMBL:U08123; NID:g473260; PIDN:AAB60444.1; PID:g473261

A;Note: correction to A37810; sequence thought by authors to be macrophage splice form

A;Accession: S45260

A;Status: translation not shown

A;Molecule type: DNA

A;Residues: 1-55 <CHO>

A;Cross-references: GB:U08122; NID:g472856; PIDN:AAA18869.1; PID:g472858;  
EMBL:U08121

R;Sato, A.; Nishino, T.; Noda, K.; Amaya, Y.; Nishino, T.  
J. Biol. Chem. 270, 2818-2826, 1995

A;Title: The structure of chicken liver xanthine dehydrogenase. cDNA cloning and the domain structure.

A;Reference number: A55711; MUID:95155354; PMID:7852355

A;Contents: annotation; confirmation of sequence

A;Note: the authors confirmed that both liver and macrophage mRNA's of the rat have a sequence in accordance with the correction in I58308

R;McManaman, J.L.; Shellman, V.; Wright, R.M.; Repine, J.E.

Arch. Biochem. Biophys. 332, 135-141, 1996

A;Title: Purification of rat liver xanthine oxidase and xanthine dehydrogenase by affinity chromatography on benzamidine-sepharose.

A;Reference number: S71397; MUID:96400342; PMID:8806718

A;Accession: S71397

A;Molecule type: protein

A;Residues: 2-11 <MCM>

C;Comment: Xanthine dehydrogenase is reversibly converted to xanthine oxidase by oxidized glutathione catalyzed by enzyme-thiol transhydrogenase (oxidized-glutathione) (EC 1.8.4.7). The reversible conversion to xanthine oxidase can also be performed artificially by a variety of sulfhydryl reagents. An irreversible conversion can be performed by limited proteolysis.

C;Comment: This enzyme contains four cofactors per subunit, one FAD, two iron-sulfur clusters, and one molybdopterin.

C;Genetics:

A;Introns: 14/3; 34/1

A;Note: the list of introns may be incomplete

C;Complex: homodimer

C;Function: <XDH>

A;Description: catalyzes oxidation of xanthine to uric acid by NAD<sup>+</sup> and water

A;Pathway: purine catabolism

C;Function: <XO>

A;Description: catalyzes oxidation of xanthine to uric acid and hydrogen peroxide by dioxygen and water

A;Pathway: purine catabolism

C;Superfamily: xanthine dehydrogenase; ferredoxin [2Fe-2S] homology

C;Keywords: 2Fe-2S; FAD; flavoprotein; homodimer; iron-sulfur protein; metalloprotein; molybdenum; molybdopterin; nucleotide binding; oxidoreductase; P-loop; peroxisome; phosphoprotein; purine catabolism

F;2-1331/Product: xanthine dehydrogenase / xanthine oxidase #status experimental <MAT>

F;26-74/Domain: ferredoxin [2Fe-2S] homology <FER1>

F;795-802/Region: nucleotide-binding motif A (P-loop)

F;43,48,51,73/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted

F;112,115,147,149/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted

F;825/Binding site: molybdopterin (Cys) (covalent) #status predicted

F;912/Binding site: molybdopterin (Arg) #status predicted

F;1261/Active site: Glu #status predicted

Query Match 83.7%; Score 36; DB 1; Length 1331;

Best Local Similarity 71.4%; Pred. No. 85;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7

|| :|||

Db 37 CGTKLGC 43

XOMSDH

xanthine dehydrogenase (EC 1.1.1.204) / xanthine oxidase (EC 1.1.3.22) - mouse

N;Alternate names: hypoxanthine oxidase

C;Species: Mus musculus (house mouse)

C;Date: 15-Mar-1996 #sequence\_revision 07-Feb-1997 #text\_change 19-Jan-2001

C;Accession: I48374; S22419; S65134

R;Cazzaniga, G.; Terao, M.; Lo Schiavo, P.; Galbiati, F.; Segalla, F.; Seldin, M.F.; Garattini, E.

Genomics 23, 390-402, 1994

A;Title: Chromosomal mapping, isolation, and characterization of the mouse xanthine dehydrogenase gene.

A;Reference number: A55561; MUID:95137585; PMID:7835888

A;Accession: I48374

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-1335 <RES>

A;Cross-references: EMBL:X75129; NID:g473040; PIDN:CAA52997.1; PID:g817959

A;Note: the sequence and translation are shown only for the splice boundaries

R;Terao, M.; Cazzaniga, G.; Ghezzi, P.; Bianchi, M.; Falciani, F.; Perani, P.; Garattini, E.

Biochem. J. 283, 863-870, 1992

A;Title: Molecular cloning of a cDNA coding for mouse liver xanthine dehydrogenase: regulation of its transcript by interferons in vivo.

A;Reference number: S22419; MUID:92272690; PMID:1590774

A;Accession: S22419

A;Molecule type: mRNA

A;Residues: 1-240, 'I', 242-620, 'M', 622-1335 <TER>

A;Cross-references: EMBL:X62932; NID:g55443; PIDN:CAA44705.1; PID:g55444

R;Ishii, T.; Aoki, N.; Noda, A.; Adachi, T.; Nakamura, R.; Matsuda, T.

Biochim. Biophys. Acta 1245, 285-292, 1995

A;Title: Carboxy-terminal cytoplasmic domain of mouse butyrophilin specifically associates with a 150-kDa protein of mammary epithelial cells and milk fat globule membrane.

A;Reference number: S65133; MUID:96125722; PMID:8541302

A;Accession: S65134

A;Molecule type: protein

A;Residues: 2-9 <ISH>

C;Comment: Xanthine dehydrogenase is reversibly converted to xanthine oxidase by oxidized glutathione catalyzed by enzyme-thiol transhydrogenase (oxidized-glutathione) (EC 1.8.4.7). The reversible conversion to xanthine oxidase can also be performed artificially by a variety of sulfhydryl reagents. An irreversible conversion can be performed by limited proteolysis.

C;Comment: This enzyme contains four cofactors per subunit, one FAD, two iron-sulfur clusters, and one molybdopterin.

C;Genetics:

A;Gene: XDH; XD; XO

A;Introns: 17/3; 37/1; 69/2; 104/3; 147/1; 167/3; 190/3; 219/3; 267/1; 298/1; 348/3; 380/1; 416/3; 478/2; 536/3; 564/3; 621/2; 662/3; 702/3; 735/1; 776/3; 821/2; 850/3; 879/3; 943/3; 992/2; 1019/3; 1051/3; 1094/3; 1119/3; 1137/2; 1175/3; 1197/3; 1260/3; 1319/3

C;Complex: homodimer

C;Function: <XDH>

A;Description: catalyzes oxidation of xanthine to uric acid by NAD<sup>+</sup> and water

A;Pathway: purine catabolism

C;Function: <XO>

A;Description: catalyzes oxidation of xanthine to uric acid and hydrogen peroxide by dioxygen and water

A;Pathway: purine catabolism  
 C;Superfamily: xanthine dehydrogenase; ferredoxin [2Fe-2S] homology  
 C;Keywords: 2Fe-2S; FAD; flavoprotein; homodimer; iron-sulfur protein;  
 metalloprotein; molybdenum; molybdopterin; nucleotide binding; oxidoreductase;  
 P-loop; peroxisome; phosphoprotein; purine catabolism  
 F;2-1335/Product: xanthine dehydrogenase / xanthine oxidase #status predicted  
 <MAT>  
 F;29-77/Domain: ferredoxin [2Fe-2S] homology <FER1>  
 F;798-805/Region: nucleotide-binding motif A (P-loop)  
 F;46,51,54,76/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted  
 F;115,118,150,152/Binding site: 2Fe-2S cluster (Cys) (covalent) #status  
 predicted  
 F;828/Binding site: molybdopterin (Cys) (covalent) #status predicted  
 F;915/Binding site: molybdopterin (Arg) #status predicted  
 F;1264/Active site: Glu #status predicted

Query Match 83.7%; Score 36; DB 1; Length 1335;  
 Best Local Similarity 71.4%; Pred. No. 86;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
 || :|||  
 Db 40 CGTKLGC 46

#### RESULT 5

S07245

xanthine dehydrogenase (EC 1.1.1.204) - fruit fly (*Drosophila melanogaster*)

C;Species: *Drosophila melanogaster*

C;Date: 21-Nov-1993 #sequence\_revision 07-Jun-1996 #text\_change 11-Jun-1999

C;Accession: S07245; S07244; S10132

R;Keith, T.P.; Riley, M.A.; Kreitman, M.; Lewontin, R.C.; Curtis, D.; Chambers, G.

Genetics 116, 67-73, 1987

A;Title: Sequence of the structural gene for xanthine dehydrogenase (rosy locus) in *Drosophila melanogaster*.

A;Reference number: S07245; MUID:87248040; PMID:3036646

A;Accession: S07245

A;Molecule type: DNA

A;Residues: 198-1335 <KEI>

A;Cross-references: EMBL:Y00308

A;Note: mRNA was also sequenced

R;Lee, C.S.; Curtis, D.; McCarron, M.; Love, C.; Gray, M.; Bender, W.; Chovnick, A.

Genetics 116, 55-66, 1987

A;Title: Mutations affecting expression of the rosy locus in *Drosophila melanogaster*.

A;Reference number: S07244; MUID:87248039; PMID:3036645

A;Accession: S07244

A;Molecule type: DNA

A;Residues: 1-230 <LEE>

A;Cross-references: EMBL:Y00308

A;Note: the authors translated the codon ACC for residue 185 as Ser

A;Note: mRNA was also sequenced

R;Lee, C.S.; Curtis, D.; McCarron, M.; Love, C.; Gray, M.; Bender, W.; Chovnick, A.

submitted to the EMBL Data Library, February 1987



A;Reference number: S10132  
 A;Accession: S10132  
 A;Molecule type: DNA  
 A;Residues: 1-105,'P',107-1335 <LE2>  
 A;Cross-references: EMBL:Y00308; NID:g8830; PIDN:CAA68409.1; PID:g8831  
 C;Genetics:  
 A;Gene: FlyBase:ry  
 A;Cross-references: FlyBase:FBgn0003308  
 A;Introns: 14/3; 881/3; 1319/3  
 C;Superfamily: xanthine dehydrogenase; ferredoxin [2Fe-2S] homology  
 C;Keywords: 2Fe-2S; FAD; flavoprotein; iron-sulfur protein; metalloprotein;  
 molybdenum; oxidoreductase; peroxisome; purine catabolism  
 F;26-74/Domain: ferredoxin [2Fe-2S] homology <FER1>  
 F;43,48,51,73/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted

Query Match 83.7%; Score 36; DB 2; Length 1335;  
 Best Local Similarity 71.4%; Pred. No. 86;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
 || :|||  
 Db 37 CGTKLGC 43

#### RESULT 6

A31946  
 xanthine dehydrogenase (EC 1.1.1.204) - fruit fly (*Drosophila pseudoobscura*)  
 C;Species: *Drosophila pseudoobscura*  
 C;Date: 22-Nov-1989 #sequence\_revision 22-Nov-1989 #text\_change 11-Jun-1999  
 C;Accession: A31946  
 R;Riley, M.A.  
 Mol. Biol. Evol. 6, 33-52, 1989  
 A;Title: Nucleotide sequence of the Xdh region in *Drosophila pseudoobscura* and  
 an analysis of the evolution of synonymous codons.  
 A;Reference number: A31946; MUID:89158785; PMID:2493563  
 A;Accession: A31946  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-1342 <RIL>  
 A;Cross-references: GB:M33977; NID:g158807; PIDN:AAA29022.1; PID:g158809  
 C;Genetics:  
 A;Gene: FlyBase:Dpse/ry  
 A;Cross-references: FlyBase:FBgn0012736  
 C;Superfamily: xanthine dehydrogenase; ferredoxin [2Fe-2S] homology  
 C;Keywords: 2Fe-2S; FAD; flavoprotein; iron-sulfur protein; metalloprotein;  
 molybdenum; oxidoreductase; peroxisome; purine catabolism  
 F;30-78/Domain: ferredoxin [2Fe-2S] homology <FER1>  
 F;47,52,55,77/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted

Query Match 83.7%; Score 36; DB 2; Length 1342;  
 Best Local Similarity 71.4%; Pred. No. 86;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
 || :|||  
 Db 41 CGTKLGC 47

# RESULT 7

JQ0407

xanthine dehydrogenase (EC 1.1.1.204) - bluebottle fly (*Calliphora vicina*)

C;Species: *Calliphora vicina*

C;Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 11-Jun-1999

C;Accession: JQ0407; A29627; S03392

R;Houde, M.; Tiveron, M.C.; Bregegere, F.

Gene 85, 391-402, 1989

A;Title: Divergence of the nucleotide sequences encoding xanthine dehydrogenase in *Calliphora vicina* and *Drosophila melanogaster*.

A;Reference number: JQ0407; MUID:90185213; PMID:2516831

A;Accession: JQ0407

A;Molecule type: DNA

A;Residues: 1-1353 <HOU>

R;Rocher-Chambonnet, C.; Berreur, P.; Houde, M.; Tiveron, M.C.; Lepesant, J.A.; Bregegere, F.

Gene 59, 201-212, 1987

A;Title: Cloning and partial characterization of the xanthine dehydrogenase gene of *Calliphora vicina*, a distant relative of *Drosophila melanogaster*.

A;Reference number: A29627; MUID:88137956; PMID:2830167

A;Accession: A29627

A;Molecule type: DNA

A;Residues: 208-367 <ROC>

A;Cross-references: GB:M18423; NID:g156143; PIDN:AAA27879.1; PID:g156144

R;Houde, M.; Tiveron, M.C.; Bregegere, F.

submitted to the EMBL Data Library, March 1988

A;Reference number: S03392

A;Accession: S03392

A;Molecule type: DNA

A;Residues: 1-387, 'F', 389-1353 <HO2>

A;Cross-references: EMBL:X07229

C;Comment: The enzyme is important in the catabolism of purines.

C;Genetics:

A;Gene: xdh

A;Introns: 27/3; 1281/3; 1337/3

C;Superfamily: xanthine dehydrogenase; ferredoxin [2Fe-2S] homology

C;Keywords: 2Fe-2S; FAD; flavoprotein; iron-sulfur protein; metalloprotein; molybdenum; oxidoreductase; peroxisome; purine catabolism

F;39-87/Domain: ferredoxin [2Fe-2S] homology <FER1>

F;56,61,64,86/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted

Query Match 83.7%; Score 36; DB 2; Length 1353;

Best Local Similarity 71.4%; Pred. No. 86;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7

|| :|||

Db 50 CGTKLGC 56

# RESULT 8

XOCHDH

xanthine dehydrogenase (EC 1.1.1.204) - chicken

C;Species: *Gallus gallus* (chicken)

C;Date: 03-Mar-1995 #sequence\_revision 07-Feb-1997 #text\_change 19-Jan-2001

C;Accession: A55711; S34758

R;Sato, A.; Nishino, T.; Noda, K.; Amaya, Y.; Nishino, T.  
 J. Biol. Chem. 270, 2818-2826, 1995  
 A;Title: The structure of chicken liver xanthine dehydrogenase. cDNA cloning and the domain structure.  
 A;Reference number: A55711; MUID:95155354; PMID:7852355  
 A;Accession: A55711  
 A;Molecule type: mRNA  
 A;Residues: 1-1358 <SAT>  
 A;Cross-references: DDBJ:D13221; NID:g507879; PIDN:BAA02502.1; PID:g507880  
 A;Note: parts of this sequence, including the amino end of the mature protein, were determined by protein sequencing  
 R;Schieber, A.; Edmondson, D.E.  
 Eur. J. Biochem. 215, 307-314, 1993  
 A;Title: Studies on the induction and phosphorylation of xanthine dehydrogenase in cultured chick embryo hepatocytes.  
 A;Reference number: S34758; MUID:93345517; PMID:8344298  
 A;Accession: S34758  
 A;Molecule type: protein  
 A;Residues: 2-20 <SCH>  
 C;Comment: This enzyme contains four cofactors per subunit, one FAD, two iron-sulfur clusters, and one molybdopterin.  
 C;Complex: homodimer  
 C;Function:  
 A;Description: catalyzes oxidation of xanthine to uric acid by NAD+ and water  
 A;Pathway: purine catabolism  
 C;Superfamily: xanthine dehydrogenase; ferredoxin [2Fe-2S] homology  
 C;Keywords: 2Fe-2S; FAD; flavoprotein; homodimer; iron-sulfur protein; metalloprotein; molybdenum; molybdopterin; nucleotide binding; oxidoreductase; P-loop; peroxisome; phosphoprotein; purine catabolism  
 F;2-1358/Product: xanthine dehydrogenase #status experimental <MAT>  
 F;30-78/Domain: ferredoxin [2Fe-2S] homology <FER1>  
 F;824-831/Region: nucleotide-binding motif A (P-loop)  
 F;47,52,55,77/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted  
 F;117,120,152,154/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted  
 F;854/Binding site: molybdopterin (Cys) (covalent) #status predicted  
 F;941/Binding site: molybdopterin (Arg) #status predicted  
 F;1290/Active site: Glu #status predicted

Query Match 83.7%; Score 36; DB 1; Length 1358;  
 Best Local Similarity 71.4%; Pred. No. 87;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
 || :|||  
 Db 41 CGTKLGC 47

RESULT 9  
 A30978  
 MSEL neurophysin-copeptin - ostrich  
 C;Species: Struthio camelus (ostrich)  
 C;Date: 17-Aug-1989 #sequence\_revision 17-Aug-1989 #text\_change 18-Jun-1993  
 C;Accession: A30978  
 R;Lazure, C.; Saayman, H.S.; Naude, R.J.; Oelofsen, W.; Chretien, M.  
 Int. J. Pept. Protein Res. 33, 46-58, 1989

A;Title: Ostrich MSEL-neurophysin belongs to the class of two-domain "big" neurophysin as indicated by complete amino acid sequence of the neurophysin/copeptin.

A;Reference number: A30978; MUID:89254272; PMID:2722398

A;Accession: A30978

A;Status: preliminary

A;Molecule type: protein

A;Residues: 1-132 <LAZ>

C;Superfamily: oxytocin-neurophysin

Query Match 79.1%; Score 34; DB 2; Length 132;  
Best Local Similarity 71.4%; Pred. No. 34;  
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
|| |||  
Db 28 CGAELGC 34

#### RESULT 10

S14480

arginine-vasotocin / neurophysin 2 precursor [validated] - chicken

N;Contains: copeptin precursor

C;Species: Gallus gallus (chicken)

C;Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 17-Nov-2000

C;Accession: S14480; B61563

R;Hunt, N.; Kluever, D.; Ivell, R.

submitted to the EMBL Data Library, November 1990

A;Description: Structure and ovarian expression of the oxytocin gene in the sheep.

A;Reference number: S14480

A;Accession: S14480

A;Molecule type: mRNA

A;Residues: 1-161 <HUN>

A;Cross-references: EMBL:X55130; NID:g62848; PIDN:CAA38923.1; PID:g62849

R;Levy, B.; Michel, G.; Chauvet, J.; Chauvet, M.T.; Acher, R.

Biosci. Rep. 7, 631-636, 1987

A;Title: Gene conversion in avian mesotocin and vasotocin genes: a recurrent mechanism linking two neurohypophyseal precursor lineages?.

A;Reference number: A61563; MUID:88108074; PMID:3427215

A;Accession: B61563

A;Status: preliminary

A;Molecule type: protein

A;Residues: 40-49;52-73 <LEV>

C;Superfamily: oxytocin-neurophysin

C;Keywords: amidated carboxyl end; glycoprotein; hormone; neuropeptide

F;1-19/Domain: signal sequence #status predicted <SIG>

F;20-28/Product: arginine-vasotocin #status predicted <VAS>

F;32-125/Product: neurophysin 2 #status experimental <NEU>

F;20-25,41-85,44-58,52-75,59-65,92-104,98-116,105-110/Disulfide bonds: #status predicted

F;28/Modified site: amidated carboxyl end (Gly) (amide in mature form from following glycine) #status predicted

F;131/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 79.1%; Score 34; DB 2; Length 161;  
Best Local Similarity 71.4%; Pred. No. 40;

Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
|| |||  
Db 59 CGAELGC 65

RESULT 11

B83583

dethiobiotin synthase PA0504 [imported] - *Pseudomonas aeruginosa* (strain PA01)

C;Species: *Pseudomonas aeruginosa*

C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 02-Aug-2002

C;Accession: B83583

R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Brinkman, F.S.L.; Hufnagle, W.O.; Kowalik, D.J.; Lagrou, M.; Garber, R.L.; Goltry, L.; Tolentino, E.; Westbrook-Wadman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, R.M.; Smith, K.A.; Spencer, D.H.; Wong, G.K.S.; Wu, Z.; Paulsen, I.T.; Reizer, J.; Saier, M.H.; Hancock, R.E.W.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A;Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen.

A;Reference number: A82950; MUID:20437337; PMID:10984043

A;Accession: B83583

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-228 <STO>

A;Cross-references: GB:AE004487; GB:AE004091; NID:g9946361; PIDN:AAG03893.1; GSPDB:GN00131; PASP:PA0504

A;Experimental source: strain PA01

C;Genetics:

A;Gene: bioD; PA0504

C;Superfamily: dethiobiotin synthetase

Query Match 79.1%; Score 34; DB 2; Length 228;  
Best Local Similarity 100.0%; Pred. No. 52;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GVRLGC 7  
|||||  
Db 147 GVRLGC 152

RESULT 12

B82694

copper homeostasis protein XF1341 [imported] - *Xylella fastidiosa* (strain 9a5c)

C;Species: *Xylella fastidiosa*

C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000

C;Accession: B82694

R;anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis, Sao Paulo, Brazil.

Nature 406, 151-157, 2000

A;Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.

A;Reference number: A82515; MUID:20365717; PMID:10910347

A;Note: for a complete list of authors see reference number A59328 below

A;Accession: B82694

A;Status: preliminary

A;Molecule type: DNA  
A;Residues: 1-267 <SIM>  
A;Cross-references: GB:AE003966; GB:AE003849; NID:g9106327; PIDN:AAF84150.1;  
GSPDB:GN00128; XFSC:XF1341  
A;Experimental source: strain 9a5c  
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.;  
Alvarenga, R.; Alves, L.M.C.; Araya, J.E.; Baia, G.S.; Baptista, C.S.; Barros,  
M.H.; Bonaccorsi, E.D.; Bordin, S.; Bove, J.M.; Briones, M.R.S.; Bueno, M.R.P.;  
Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H.; Colauto, N.B.;  
Colombo, C.; Costa, F.F.; Costa, M.C.R.; Costa-Neto, C.M.; Coutinho, L.L.;  
Cristofani, M.; Dias-Neto, E.; Docena, C.; El-Dorrry, H.; Facincani, A.P.;  
Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco,  
M.C.; Frohme, M.; Furlan, L.R.; Garnier, M.; Goldman, G.H.; Goldman, M.H.S.;  
Gomes, S.L.; Gruber, A.; Ho, P.L.; Hoheisel, J.D.; Junqueira, M.L.; Kemper,  
E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigret, F.; Lambais, M.R.;  
Leite, L.C.C.; Lemos, E.G.M.; Lemos, M.V.F.; Lopes, S.A.; Lopes, C.R.; Machado,  
J.A.; Machado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques,  
M.V.; Martins, E.A.L.  
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.;  
Miyaki, C.Y.; Monteiro-Vitorello, C.B.; Moon, D.H.; Nagai, M.A.; Nascimento,  
A.L.T.O.; Netto, L.E.S.; Nhani Jr., A.; Nobrega, F.G.; Nunes, L.R.; Oliveira,  
M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.; Paris, A.; Peixoto,  
B.R.; Pereira, G.A.G.; Pereira Jr., H.A.; Pesquero, J.B.; Quaggio, R.B.;  
Roberto, P.G.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.;  
Santelli, R.V.; Sawasaki, H.E.  
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da  
Silveira, J.F.; Silvestri, M.L.Z.; Siqueira, W.J.; de Souza, A.A.; de Souza,  
A.P.; Terenzi, M.F.; Truffi, D.; Tsai, S.M.; Tsuhako, M.H.; Vallada, H.; Van  
Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Zago, M.A.; Zatz, M.;  
Meidanis, J.; Setubal, J.C.  
A;Reference number: A59328  
A;Contents: annotation  
C;Genetics:  
A;Gene: XF1341

Query Match 79.1%; Score 34; DB 2; Length 267;  
Best Local Similarity 85.7%; Pred. No. 59;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
| |||||  
Db 105 CCVRLGC 111

RESULT 13  
T01943  
hypothetical protein F1104.2 - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 26-Feb-1999 #sequence\_revision 26-Feb-1999 #text\_change 24-Mar-1999  
C;Accession: T01943  
R;Abu-Threideh, J.; Stoneking, T.; Langston, Y.; Trevaskis, E.  
submitted to the EMBL Data Library, October 1998  
A;Description: The sequence of A. thaliana F1104.  
A;Reference number: Z14466  
A;Accession: T01943

A;Status: translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-382 <ABU>  
A;Cross-references: EMBL:AF096370; NID:g3695372; PID:g3695376  
A;Experimental source: cultivar Columbia  
C;Genetics:  
A;Map position: 4  
A;Introns: 163/1  
A;Note: F1104.2

Query Match 79.1%; Score 34; DB 2; Length 382;  
Best Local Similarity 71.4%; Pred. No. 77;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
||:| ||  
Db 104 CGLRQGC 110

#### RESULT 14

S37855

hypothetical protein YKL034w precursor - yeast (*Saccharomyces cerevisiae*)

N;Alternate names: hypothetical protein YKL247

C;Species: *Saccharomyces cerevisiae*

C;Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 19-Apr-2002

C;Accession: S37855; S41670; S36853

R;Purnelle, B.; Skala, J.; van Dyck, L.; Tettelin, H.; Goffeau, A.

submitted to the Protein Sequence Database, March 1994

A;Reference number: S37851

A;Accession: S37855

A;Molecule type: DNA

A;Residues: 1-758 <PUR>

A;Cross-references: EMBL:Z28034; NID:g486043; PIDN:CAA81869.1; PID:g486044;

MIPS:YKL034w

A;Experimental source: strain S288C

R;Purnelle, B.; Skala, J.; van Dyck, L.; Goffeau, A.

Yeast 10, 125-130, 1994

A;Title: Analysis of an 11.7 kb DNA fragment of chromosome XI reveals a new tRNA gene and four new open reading frames including a leucine zipper protein and a homologue to the yeast mitochondrial regulator ABF2.

A;Reference number: S41667; MUID:94262309; PMID:8203146

A;Accession: S41670

A;Status: nucleic acid sequence not shown

A;Molecule type: DNA

A;Residues: 1-758 <PU2>

A;Cross-references: EMBL:X71622

A;Experimental source: strain S288C

R;Purnelle, B.; Skala, J.; van Dyck, L.; Goffeau, A.

Yeast 8, 977-986, 1992

A;Title: The sequence of a 12 kb fragment on the left arm of yeast chromosome XI reveals five new open reading frames, including a zinc finger protein and a homolog of the UDP-glucose pyrophosphorylase from potato.

A;Reference number: S30007; MUID:93127731; PMID:1481573

A;Accession: S36853

A;Status: translation not shown

A;Molecule type: DNA

A;Residues: 1-570 <PU3>

A;Cross-references: EMBL:X69584; NID:g4789; PIDN:CAA49298.1; PID:g871537  
A;Experimental source: strain S288C  
C;Genetics:  
A;Cross-references: SGD:S0001517  
A;Map position: 11L  
C;Superfamily: Saccharomyces cerevisiae hypothetical protein YKL034w; RING  
finger homology  
C;Keywords: transmembrane protein  
F;1-26/Domain: signal sequence #status predicted <SIG>  
F;27-758/Product: hypothetical protein YKL034w #status predicted <MAT>  
F;400-416/Domain: transmembrane #status predicted <TM1>  
F;440-456/Domain: transmembrane #status predicted <TM2>  
F;461-477/Domain: transmembrane #status predicted <TM3>  
F;528-544/Domain: transmembrane #status predicted <TM4>  
F;607-623/Domain: transmembrane #status predicted <TM5>  
F;638-654/Domain: transmembrane #status predicted <TM6>  
F;695-757/Domain: RING finger homology <RRN>

Query Match 79.1%; Score 34; DB 2; Length 758;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGVRLG 6  
|||||  
Db 375 CGVRLG 380

# RESULT 15

T02805  
chloride channel protein CCP [imported] - Leishmania major (strain Friedlin)  
C;Species: Leishmania major  
C;Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 19-May-2000  
C;Accession: A81457; T02805  
R;Myler, P.J.; Audleman, L.; deVos, T.; Hixson, G.; Kiser, P.; Lemley, C.;  
Magness, C.; Rickel, E.; Sisk, E.; Sunkin, S.; Swartzell, S.; Westlake, T.;  
Bastien, P.; Fu, G.; Ivens, A.; Stuart, K.  
Proc. Natl. Acad. Sci. U.S.A. 96, 2902-2906, 1999  
A;Title: Leishmania major Friedlin chromosome 1 has an unusual distribution of  
protein-coding genes.  
A;Reference number: A81455; MUID:99178987; PMID:10077609  
A;Accession: A81457  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-772 <PYL>  
A;Cross-references: GB:AE001274; NID:g3264850; PIDN:AAC24628.1; PID:g2995581;  
GSPDB:GN00125  
A;Experimental source: strain MHOM/IL/81/Friedlin  
C;Genetics:  
A;Gene: CCP  
A;Map position: 1

Query Match 79.1%; Score 34; DB 2; Length 772;  
Best Local Similarity 71.4%; Pred. No. 1.3e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
||: |||



Db 94 CGIVLGC 100

Search completed: November 13, 2003, 09:52:57  
Job time : 9.29167 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:31:40 ; Search time 4.01042 Seconds  
(without alignments)  
82.083 Million cell updates/sec

Title: US-09-228-866-6  
Perfect score: 43  
Sequence: 1 CGVRLGC 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result		%					
No.	Score	Query	Match	Length	DB	ID	Description
1	36	83.7	438	1	IAP1_DROME	Q24306	drosophila
2	36	83.7	662	1	SL51_RABIT	P11170	oryctolagus
3	36	83.7	1330	1	XDH_RAT	P22985	rattus norv
4	36	83.7	1335	1	XDH_DROME	P10351	drosophila
5	36	83.7	1335	1	XDH_MOUSE	Q00519	mus musculu
6	36	83.7	1342	1	XDH_DROPS	P22811	drosophila
7	36	83.7	1344	1	XDH_DROSU	P91711	drosophila
8	36	83.7	1353	1	XDH_CALVI	P08793	calliphora
9	36	83.7	1358	1	XDH_CHICK	P47990	gallus gall
10	34	79.1	132	1	NEU2_STRCA	P21916	struthio ca
11	34	79.1	161	1	NEUV_CHICK	P24787	gallus gall
12	34	79.1	224	1	BIOD_XANAC	Q8pgk0	xanthomonas

13	34	79.1	224	1	BIOD_XANCP	Q8pcw4	xanthomonas
14	34	79.1	228	1	BIOD_PSEAE	Q9i614	pseudomonas
15	34	79.1	404	1	KVB3_HUMAN	O43448	homo sapien
16	34	79.1	758	1	YKD4_YEAST	P36096	saccharomyc
17	34	79.1	1402	1	IF4G_RABIT	P41110	oryctolagus
18	33	76.7	152	1	NEU3_CATCO	P17668	catostomus
19	33	76.7	184	1	MAA_BACSU	P37515	bacillus su
20	33	76.7	269	1	CYSQ_ACTAC	P70714	actinobacil
21	33	76.7	605	1	SL51_PIG	P26429	sus scrofa
22	33	76.7	656	1	SL54_MOUSE	Q9et37	mus musculu
23	33	76.7	659	1	SL54_HUMAN	Q9ny91	homo sapien
24	33	76.7	664	1	SL51_HUMAN	P13866	homo sapien
25	33	76.7	664	1	SL51_SHEEP	P53791	ovis aries
26	32	74.4	79	1	PSPB_BOVIN	P15781	bos taurus
27	32	74.4	92	1	NEU2_LOXAF	P81768	loxodonta a
28	32	74.4	93	1	NEU1_ANSAN	P35519	anser anser
29	32	74.4	93	1	NEU1_STRCA	P15444	struthio ca
30	32	74.4	122	1	PA22_VIPAZ	Q10755	vipera aspi
31	32	74.4	125	1	NEUM_BUFJA	P08162	bufo japoni
32	32	74.4	327	1	Y349_CHLPN	Q9z8j4	chlamydia p
33	32	74.4	395	1	ARGJ_METAC	Q8tk55	m arginine
34	32	74.4	402	1	ARGJ_METMA	Q8pzl8	m arginine
35	32	74.4	605	1	SYA_TREPA	O83980	treponema p
36	32	74.4	718	1	SL53_BOVIN	P53793	bos taurus
37	32	74.4	718	1	SL53_CANFA	P31637	canis famil
38	32	74.4	718	1	SL53_HUMAN	P53794	homo sapien
39	32	74.4	718	1	SL53_MOUSE	Q9jkz2	mus musculu
40	31	72.1	92	1	NEU2_HORSE	P01182	equus cabal
41	31	72.1	105	1	NEU1_HORSE	P01176	equus cabal
42	31	72.1	107	1	NEU2_BALPH	P01184	balaenopter
43	31	72.1	108	1	Y14A_BPT4	P32279	bacterioph
44	31	72.1	125	1	NEU1_BOVIN	P01175	bos taurus
45	31	72.1	125	1	NEU1_PIG	P01177	sus scrofa

# ALIGNMENTS

## RESULT 1

IAP1\_DROME

ID IAP1\_DROME STANDARD; PRT; 438 AA.

AC Q24306;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Apoptosis 1 inhibitor (Inhibitor of apoptosis 1) (dIAP1) (Thread protein).

GN IAP1 OR TH.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Eye imaginal disk;

RX MEDLINE=96128128; PubMed=8548811;

RA Hay B.A., Wassarman D.A., Rubin G.M.;  
 RT "Drosophila homologs of baculovirus inhibitor of apoptosis proteins  
 RT function to block cell death.";  
 RL Cell 83:1253-1262(1995).  
 CC -!- FUNCTION: APOPTOTIC SUPPRESSOR. OVEREXPRESSION SUPPRESSES RPR AND  
 CC HID-DEPENDENT CELL DEATH IN THE EYE.  
 CC -!- SIMILARITY: BELONGS TO THE IAP FAMILY.  
 CC -!- SIMILARITY: Contains 2 BIR repeats.  
 CC -!- SIMILARITY: Contains 1 RING-type zinc finger.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L49440; AAC41609.1; -.  
 DR PDB; 1JD4; 05-DEC-01.  
 DR PDB; 1JD5; 05-DEC-01.  
 DR PDB; 1JD6; 05-DEC-01.  
 DR FlyBase; FBgn0003691; th.  
 DR GO; GO:0008189; F:apoptosis inhibitor activity; IDA.  
 DR GO; GO:0006916; P:anti-apoptosis; IDA.  
 DR InterPro; IPR001370; BIR.  
 DR InterPro; IPR001841; Znf\_ring.  
 DR Pfam; PF00653; BIR; 2.  
 DR Pfam; PF00097; zf-C3HC4; 1.  
 DR SMART; SM00238; BIR; 2.  
 DR SMART; SM00184; RING; 1.  
 DR PROSITE; PS01282; BIR\_REPEAT\_1; 2.  
 DR PROSITE; PS50143; BIR\_REPEAT\_2; 2.  
 DR PROSITE; PS00518; ZF\_RING\_1; FALSE\_NEG.  
 DR PROSITE; PS50089; ZF\_RING\_2; 1.  
 KW Apoptosis; Zinc-finger; Repeat; 3D-structure.  
 FT REPEAT 44 110 BIR 1.  
 FT REPEAT 226 293 BIR 2.  
 FT ZN\_FING 391 426 RING-TYPE.  
 SQ SEQUENCE 438 AA; 48098 MW; A6C22C8EDF5AEF29 CRC64;

Query Match 83.7%; Score 36; DB 1; Length 438;  
 Best Local Similarity 71.4%; Pred. No. 11;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
 ||| :||  
 Db 83 CGVEIGC 89

## RESULT 2

### SL51\_RABIT

ID SL51\_RABIT STANDARD; PRT; 662 AA.  
 AC P11170;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE Sodium/glucose cotransporter 1 (Na(+)/glucose cotransporter 1)  
 DE (High affinity sodium-glucose cotransporter).  
 GN SLC5A1 OR SGLT1.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=New Zealand white;  
 RX MEDLINE=88065856; PubMed=2446136;  
 RA Hediger M.A., Coady M.J., Ikeda T.S., Wright E.M.;  
 RT "Expression cloning and cDNA sequencing of the Na<sup>+</sup>/glucose co-  
 RT transporter.";  
 RL Nature 330:379-381(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=New Zealand white; TISSUE=Kidney cortex;  
 RX MEDLINE=91223090; PubMed=2025641;  
 RA Morrison A.I., Panayotova-Heiermann M., Feigl G., Schoelermann B.,  
 RA Kinne R.K.H.;  
 RT "Sequence comparison of the sodium-D-glucose cotransport systems in  
 RT rabbit renal and intestinal epithelia.";  
 RL Biochim. Biophys. Acta 1089:121-123(1991).  
 CC -!- FUNCTION: ACTIVELY TRANSPORTS GLUCOSE INTO CELLS BY NA<sup>+</sup>  
 CC CO-TRANSPORT WITH A NA<sup>+</sup> TO GLUCOSE COUPLING RATIO OF 2:1.  
 CC -!- FUNCTION: EFFICIENT SUBSTRATE TRANSPORT IN MAMMALIAN KIDNEY IS  
 CC PROVIDED BY THE CONCERTED ACTION OF A LOW AFFINITY HIGH CAPACITY  
 CC AND A HIGH AFFINITY LOW CAPACITY NA(+)/GLUCOSE COTRANSPORTER  
 CC ARRANGED IN SERIES ALONG KIDNEY PROXIMAL TUBULES.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- TISSUE SPECIFICITY: FOUND PREDOMINANTLY IN INTESTINE AND IN OUTER  
 CC RENAL MEDULLA.  
 CC -!- DISEASE: MUTATION OF ASP-28 IS IMPLICATED IN GLUCOSE/GALACTOSE  
 CC MALABSORPTION.  
 CC -!- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X06419; CAA29727.1; -.  
 DR EMBL; X55355; CAA39040.1; -.  
 DR PIR; S00515; A37226.  
 DR InterPro; IPR001734; Na/solut\_sympoort.  
 DR Pfam; PF00474; SSF; 1.  
 DR TIGRFAMs; TIGR00813; sss; 1.  
 DR PROSITE; PS00456; NA\_SOLUT\_SYMP\_1; 1.  
 DR PROSITE; PS00457; NA\_SOLUT\_SYMP\_2; 1.  
 DR PROSITE; PS50283; NA\_SOLUT\_SYMP\_3; 1.  
 KW Transport; Sugar transport; Transmembrane; Sodium transport; Symport;  
 KW Glycoprotein.  
 FT DOMAIN 1 28 CYTOPLASMIC (POTENTIAL).

FT	TRANSMEM	29	47	POTENTIAL.
FT	DOMAIN	48	64	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	65	85	POTENTIAL.
FT	DOMAIN	86	105	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	106	126	POTENTIAL.
FT	DOMAIN	127	171	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	172	191	POTENTIAL.
FT	DOMAIN	192	208	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	209	229	POTENTIAL.
FT	DOMAIN	230	270	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	271	291	POTENTIAL.
FT	DOMAIN	292	314	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	315	334	POTENTIAL.
FT	DOMAIN	335	423	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	424	443	POTENTIAL.
FT	DOMAIN	444	455	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	456	476	POTENTIAL.
FT	DOMAIN	477	526	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	527	547	POTENTIAL.
FT	DOMAIN	548	640	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	641	661	POTENTIAL.
FT	DOMAIN	662	662	EXTRACELLULAR (POTENTIAL).
FT	CARBOHYD	248	248	N-LINKED (GLCNAC. . .).
FT	SITE	43	43	IMPLICATED IN SODIUM COUPLING
FT				(BY SIMILARITY).
FT	SITE	300	300	IMPLICATED IN SODIUM COUPLING
FT				(BY SIMILARITY).
SQ	SEQUENCE	662 AA; 73079 MW; 03F55A0309CBBE01 CRC64;		

Query Match 83.7%; Score 36; DB 1; Length 662;  
 Best Local Similarity 71.4%; Pred. No. 16;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
 || |:||  
 Db 355 CGTRVGC 361

### RESULT 3

XDH\_RAT  
 ID XDH\_RAT STANDARD; PRT; 1330 AA.  
 AC P22985; Q63157;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Xanthine dehydrogenase/oxidase [Includes: Xanthine dehydrogenase  
 DE (EC 1.1.1.204) (XD); Xanthine oxidase (EC 1.1.3.22) (XO) (Xanthine  
 DE oxidoreductase)].  
 GN XDH.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Liver;  
 RX MEDLINE=90354396; PubMed=2387845;

RA Amaya Y., Yamazaki K.-I., Sato M., Noda K., Nishino T., Nishino T.;  
 RT "Proteolytic conversion of xanthine dehydrogenase from the  
 RT NAD-dependent type to the O<sub>2</sub>-dependent type. Amino acid sequence of  
 RT rat liver xanthine dehydrogenase and identification of the cleavage  
 RT sites of the enzyme protein during irreversible conversion by  
 RT trypsin.";  
 RL J. Biol. Chem. 265:14170-14175(1990).  
 RN [2]  
 RP SEQUENCE OF 1-54 FROM N.A.  
 RC STRAIN=Sprague-Dawley;  
 RX MEDLINE=94268906; PubMed=8208609;  
 RA Chow C.W., Clark M., Rinaldo J., Chalkley R.;  
 RT "Identification of the rat xanthine dehydrogenase/oxidase promoter.";  
 RL Nucleic Acids Res. 22:1846-1854(1994).  
 CC -!- FUNCTION: THIS ENZYME CAN BE CONVERTED FROM THE DEHYDROGENASE FORM  
 CC (D) TO THE OXIDASE FORM (O) IRREVERSIBLY BY PROTEOLYSIS OR  
 CC REVERSIBLY THROUGH THE OXIDATION OF SULFHYDRYL GROUPS.  
 CC -!- CATALYTIC ACTIVITY: Xanthine + NAD(+) + H(2)O = urate + NADH.  
 CC -!- CATALYTIC ACTIVITY: Xanthine + H(2)O + O(2) = urate + H(2)O(2).  
 CC -!- COFACTOR: FAD, MOLYBDOPTERIN, AND TWO 2FE-2S CLUSTERS.  
 CC -!- SUBUNIT: Homodimer.  
 CC -!- SUBCELLULAR LOCATION: Peroxisomal.  
 CC -!- INDUCTION: By interferon.  
 CC -!- SIMILARITY: BELONGS TO THE XANTHINE DEHYDROGENASE FAMILY.  
 CC -!- SIMILARITY: TO 2FE-2S FERREDOXINS IN THE N-TERMINAL DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; J05579; AAA42349.1; -.  
 DR EMBL; U08122; AAA18869.1; -.  
 DR EMBL; U08120; AAA18869.1; JOINED.  
 DR EMBL; U08121; AAA18869.1; JOINED.  
 DR HSSP; P80457; 1FIQ.  
 DR InterPro; IPR002888; 2Fe-2S\_bind.  
 DR InterPro; IPR006058; 2Fe2S\_ferredoxin.  
 DR InterPro; IPR000674; Aldxan\_dh\_C.  
 DR InterPro; IPR005107; CO\_deh\_flav\_C.  
 DR InterPro; IPR002346; dehydrog\_molyb.  
 DR InterPro; IPR000572; Euk\_Mb\_oxred.  
 DR InterPro; IPR001041; Ferredoxin.  
 DR Pfam; PF02738; Ald\_Xan\_dh\_C2; 1.  
 DR Pfam; PF01315; Ald\_Xan\_dh\_C; 1.  
 DR Pfam; PF03450; CO\_deh\_flav\_C; 1.  
 DR Pfam; PF00941; FAD\_binding\_5; 1.  
 DR Pfam; PF00111; fer2; 1.  
 DR Pfam; PF01799; fer2\_2; 1.  
 DR ProDom; PD186071; 2Fe-2S\_bind; 1.  
 DR PROSITE; PS00197; 2FE2S\_FERREDOXIN; 1.  
 DR PROSITE; PS00559; MOLYBDOPTERIN\_EUK; 1.  
 KW Oxidoreductase; NAD; Molybdenum; Flavoprotein; FAD; Metal-binding;  
 KW Iron-sulfur; Iron; 2Fe-2S.

```

FT  INIT_MET      0      0
FT  METAL         36     36      IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT  METAL         42     42      IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT  METAL         47     47      IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT  METAL         50     50      IRON-SULFUR (2FE-2S) (BY SIMILARITY).
SQ  SEQUENCE     1330 AA;  146111 MW;  A3DD206B9D74E565 CRC64;

  Query Match          83.7%;  Score 36;  DB 1;  Length 1330;
  Best Local Similarity 71.4%;  Pred. No. 30;
  Matches      5;  Conservative      1;  Mismatches      1;  Indels      0;  Gaps      0;

Qy      1 CGVRLGC 7
      || :|||
Db      36 CGTKLGC 42

```

#### RESULT 4

```

XDH_DROME
ID  XDH_DROME      STANDARD;      PRT;  1335 AA.
AC  P10351;
DT  01-MAR-1989 (Rel. 10, Created)
DT  01-MAR-1989 (Rel. 10, Last sequence update)
DT  28-FEB-2003 (Rel. 41, Last annotation update)
DE  Xanthine dehydrogenase (EC 1.1.1.204) (XD) (Rosy locus protein).
GN  RY OR XDH.
OS  Drosophila melanogaster (Fruit fly).
OC  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC  Ephydroidea; Drosophilidae; Drosophila.
OX  NCBI_TaxID=7227;
RN  [1]
RP  SEQUENCE OF 1-231 FROM N.A.
RC  STRAIN=Canton-S;
RX  MEDLINE=87248039; PubMed=3036645;
RA  Lee C.S., Curtis D., Gray M., Bender W.;
RT  "Mutations affecting expression of the rosy locus in Drosophila
RT  melanogaster.";
RL  Genetics 116:55-66(1987).
RN  [2]
RP  SEQUENCE OF 199-1335 FROM N.A.
RC  STRAIN=Canton-S;
RX  MEDLINE=87248040; PubMed=3036646;
RA  Keith T.P., Riley M.A., Kreitman M., Lewontin R.C., Curtis D.,
RA  Chambers G.;
RT  "Sequence of the structural gene for xanthine dehydrogenase (rosy
RT  locus) in Drosophila melanogaster.";
RL  Genetics 116:67-73(1987).
RN  [3]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Canton-S;
RA  Riley M.;
RL  Submitted (FEB-1987) to the EMBL/GenBank/DDBJ databases.
CC  -!- CATALYTIC ACTIVITY: Xanthine + NAD(+) + H(2)O = urate + NADH.
CC  -!- COFACTOR: FAD, MOLYBDOPTERIN, AND TWO 2FE-2S CLUSTERS.
CC  -!- SUBUNIT: Homodimer.
CC  -!- SUBCELLULAR LOCATION: Peroxisomal.
CC  -!- SIMILARITY: BELONGS TO THE XANTHINE DEHYDROGENASE FAMILY.

```

```

CC      -!- SIMILARITY: TO 2FE-2S FERREDOXINS IN THE N-TERMINAL DOMAIN.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; Y00308; CAA68409.1; -.
DR      PIR; S07245; S07245.
DR      HSSP; P80457; 1FO4.
DR      FlyBase; FBgn0003308; ry.
DR      InterPro; IPR002888; 2Fe-2S_bind.
DR      InterPro; IPR006058; 2Fe2S_ferredoxin.
DR      InterPro; IPR000674; Aldxan_dh_C.
DR      InterPro; IPR005107; CO_deh_flav_C.
DR      InterPro; IPR002346; dehydrog_molyb.
DR      InterPro; IPR000572; Euk_Mb_oxred.
DR      InterPro; IPR001041; Ferredoxin.
DR      Pfam; PF02738; Ald_Xan_dh_C2; 1.
DR      Pfam; PF01315; Ald_Xan_dh_C; 1.
DR      Pfam; PF03450; CO_deh_flav_C; 1.
DR      Pfam; PF00941; FAD_binding_5; 1.
DR      Pfam; PF00111; fer2; 1.
DR      Pfam; PF01799; fer2_2; 1.
DR      ProDom; PD186071; 2Fe-2S_bind; 1.
DR      PROSITE; PS00197; 2FE2S_FERREDOXIN; 1.
DR      PROSITE; PS00559; MOLYBDOPTERIN_EUK; 1.
KW      Oxidoreductase; NAD; Molybdenum; Flavoprotein; FAD; Metal-binding;
KW      Iron-sulfur; Iron; 2Fe-2S; Peroxisome.
FT      METAL          37      37      IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT      METAL          43      43      IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT      METAL          48      48      IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT      METAL          51      51      IRON-SULFUR (2FE-2S) (BY SIMILARITY).
SQ      SEQUENCE      1335 AA; 146925 MW; B37C5F4393035689 CRC64;

```

```

Query Match          83.7%; Score 36; DB 1; Length 1335;
Best Local Similarity 71.4%; Pred. No. 30;
Matches      5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 CGVRLGC 7
        || :|||
Db      37 CGTKLGC 43

```

# RESULT 5

## XDH\_MOUSE

```

ID      XDH_MOUSE      STANDARD;      PRT; 1335 AA.
AC      Q00519;
DT      01-DEC-1992 (Rel. 24, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Xanthine dehydrogenase/oxidase [Includes: Xanthine dehydrogenase
DE      (EC 1.1.1.204) (XD); Xanthine oxidase (EC 1.1.3.22) (XO) (Xanthine
DE      oxidoreductase)].

```



GN XDH.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129/Sv; TISSUE=Spleen;  
 RX MEDLINE=95137585; PubMed=7835888;  
 RA Cazzaniga G., Terao M., Lo Schiavo P., Galbiati F., Segalla F.,  
 RA Seldin M.F., Garattini E.;  
 RT "Chromosomal mapping, isolation, and characterization of the mouse  
 RT xanthine dehydrogenase gene.";  
 RL Genomics 23:390-402(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Liver;  
 RX MEDLINE=92272690; PubMed=1590774;  
 RA Terao M., Cazzaniga G., Ghezzi P., Bianchi M., Falciani F.,  
 RA Perani P., Garattini E.;  
 RT "Molecular cloning of a cDNA coding for mouse liver xanthine  
 RT dehydrogenase. Regulation of its transcript by interferons in vivo.";  
 RL Biochem. J. 283:863-870(1992).  
 CC -!- FUNCTION: THIS ENZYME CAN BE CONVERTED FROM THE DEHYDROGENASE FORM  
 CC (D) TO THE OXIDASE FORM (O) IRREVERSIBLY BY PROTEOLYSIS OR  
 CC REVERSIBLY THROUGH THE OXIDATION OF SULFHYDRYL GROUPS.  
 CC -!- CATALYTIC ACTIVITY: Xanthine + NAD(+) + H(2)O = urate + NADH.  
 CC -!- CATALYTIC ACTIVITY: Xanthine + H(2)O + O(2) = urate + H(2)O(2).  
 CC -!- COFACTOR: FAD, MOLYBDOPTERIN, AND TWO 2FE-2S CLUSTERS.  
 CC -!- SUBUNIT: Homodimer.  
 CC -!- SUBCELLULAR LOCATION: Peroxisomal.  
 CC -!- INDUCTION: By interferon.  
 CC -!- SIMILARITY: BELONGS TO THE XANTHINE DEHYDROGENASE FAMILY.  
 CC -!- SIMILARITY: TO 2FE-2S FERREDOXINS IN THE N-TERMINAL DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X75129; CAA52997.1; -.  
 DR EMBL; X75128; CAA52997.1; JOINED.  
 DR EMBL; X75127; CAA52997.1; JOINED.  
 DR EMBL; X75126; CAA52997.1; JOINED.  
 DR EMBL; X75125; CAA52997.1; JOINED.  
 DR EMBL; X75124; CAA52997.1; JOINED.  
 DR EMBL; X75123; CAA52997.1; JOINED.  
 DR EMBL; X75122; CAA52997.1; JOINED.  
 DR EMBL; X75121; CAA52997.1; JOINED.  
 DR EMBL; X75120; CAA52997.1; JOINED.  
 DR EMBL; X75119; CAA52997.1; JOINED.  
 DR EMBL; X75130; CAA52997.1; JOINED.  
 DR EMBL; X75131; CAA52997.1; JOINED.  
 DR EMBL; X75132; CAA52997.1; JOINED.

DR EMBL; X75133; CAA52997.1; JOINED.  
 DR EMBL; X75134; CAA52997.1; JOINED.  
 DR EMBL; X75135; CAA52997.1; JOINED.  
 DR EMBL; X75136; CAA52997.1; JOINED.  
 DR EMBL; X75137; CAA52997.1; JOINED.  
 DR EMBL; X75138; CAA52997.1; JOINED.  
 DR EMBL; X75139; CAA52997.1; JOINED.  
 DR EMBL; X75140; CAA52997.1; JOINED.  
 DR EMBL; X75141; CAA52997.1; JOINED.  
 DR EMBL; X75142; CAA52997.1; JOINED.  
 DR EMBL; X75143; CAA52997.1; JOINED.  
 DR EMBL; X75151; CAA52997.1; JOINED.  
 DR EMBL; X75152; CAA52997.1; JOINED.  
 DR EMBL; X75153; CAA52997.1; JOINED.  
 DR EMBL; X75154; CAA52997.1; JOINED.  
 DR EMBL; X75144; CAA52997.1; JOINED.  
 DR EMBL; X75145; CAA52997.1; JOINED.  
 DR EMBL; X75146; CAA52997.1; JOINED.  
 DR EMBL; X75147; CAA52997.1; JOINED.  
 DR EMBL; X75148; CAA52997.1; JOINED.  
 DR EMBL; X75149; CAA52997.1; JOINED.  
 DR EMBL; X75150; CAA52997.1; JOINED.  
 DR EMBL; X62932; CAA44705.1; -.  
 DR PIR; I48374; KOMSDH.  
 DR HSSP; P80457; 1FO4.  
 DR MGD; MGI:98973; Xdh.  
 DR InterPro; IPR002888; 2Fe-2S\_bind.  
 DR InterPro; IPR006058; 2Fe2S\_ferredoxin.  
 DR InterPro; IPR000674; Aldxan\_dh\_C.  
 DR InterPro; IPR005107; CO\_deh\_flav\_C.  
 DR InterPro; IPR002346; dehydrog\_molyb.  
 DR InterPro; IPR000572; Euk\_Mb\_oxred.  
 DR InterPro; IPR001041; Ferredoxin.  
 DR Pfam; PF02738; Ald\_Xan\_dh\_C2; 1.  
 DR Pfam; PF01315; Ald\_Xan\_dh\_C; 1.  
 DR Pfam; PF03450; CO\_deh\_flav\_C; 1.  
 DR Pfam; PF00941; FAD\_binding\_5; 1.  
 DR Pfam; PF00111; fer2; 1.  
 DR Pfam; PF01799; fer2\_2; 1.  
 DR ProDom; PD186071; 2Fe-2S\_bind; 1.  
 DR PROSITE; PS00197; 2FE2S\_FERREDOXIN; 1.  
 DR PROSITE; PS00559; MOLYBDOPTERIN\_EUK; 1.  
 KW Oxidoreductase; NAD; Molybdenum; Flavoprotein; FAD; Metal-binding;  
 KW Iron-sulfur; Iron; 2Fe-2S.  
 FT METAL 40 40 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 FT METAL 46 46 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 FT METAL 51 51 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 FT METAL 54 54 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 FT CONFLICT 241 241 V -> I (IN REF. 2).  
 FT CONFLICT 621 621 T -> M (IN REF. 2).  
 SQ SEQUENCE 1335 AA; 146517 MW; 99CE6FD8B42FB5E5 CRC64;

Query Match 83.7%; Score 36; DB 1; Length 1335;  
 Best Local Similarity 71.4%; Pred. No. 30;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGVRLGC 7

Db

|| : |||  
40 CGTKLGC 46

RESULT 6

XDH\_DROPS

ID XDH\_DROPS STANDARD; PRT; 1342 AA.  
AC P22811;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Xanthine dehydrogenase (EC 1.1.1.204) (XD) (Rosy locus protein).  
GN RY OR XDH.  
OS *Drosophila pseudoobscura* (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; *Drosophila*.  
OX NCBI\_TaxID=7237;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89158785; PubMed=2493563;  
RA Riley M.A.;  
RT "Nucleotide sequence of the Xdh region in *Drosophila pseudoobscura*  
RT and an analysis of the evolution of synonymous codons.";  
RL Mol. Biol. Evol. 6:33-52(1989).  
CC -!- CATALYTIC ACTIVITY: Xanthine + NAD(+) + H(2)O = urate + NADH.  
CC -!- COFACTOR: FAD, MOLYBDOPTERIN, AND TWO 2FE-2S CLUSTERS.  
CC -!- SUBUNIT: Homodimer.  
CC -!- SUBCELLULAR LOCATION: Peroxisomal.  
CC -!- SIMILARITY: BELONGS TO THE XANTHINE DEHYDROGENASE FAMILY.  
CC -!- SIMILARITY: TO 2FE-2S FERREDOXINS IN THE N-TERMINAL DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M33977; AAA29022.1; -.  
DR PIR; A31946; A31946.  
DR HSSP; P80457; 1FO4.  
DR FlyBase; FBgn0012736; Dpse\ry.  
DR InterPro; IPR002888; 2Fe-2S\_bind.  
DR InterPro; IPR006058; 2Fe2S\_ferredoxin.  
DR InterPro; IPR000674; Aldxan\_dh\_C.  
DR InterPro; IPR005107; CO\_deh\_flav\_C.  
DR InterPro; IPR002346; dehydrog\_molyb.  
DR InterPro; IPR000572; Euk\_Mb\_oxred.  
DR InterPro; IPR001041; Ferredoxin.  
DR Pfam; PF02738; Ald\_Xan\_dh\_C2; 1.  
DR Pfam; PF01315; Ald\_Xan\_dh\_C; 1.  
DR Pfam; PF03450; CO\_deh\_flav\_C; 1.  
DR Pfam; PF00941; FAD\_binding\_5; 1.  
DR Pfam; PF00111; fer2; 1.  
DR Pfam; PF01799; fer2\_2; 1.

DR ProDom; PD186071; 2Fe-2S\_bind; 1.  
 DR PROSITE; PS00197; 2FE2S\_FERREDOXIN; 1.  
 DR PROSITE; PS00559; MOLYBDOPTERIN\_EUK; 1.  
 KW Oxidoreductase; NAD; Molybdenum; Flavoprotein; FAD; Metal-binding;  
 KW Iron-sulfur; Iron; 2Fe-2S; Peroxisome.  
 FT METAL 41 41 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 FT METAL 47 47 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 FT METAL 52 52 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 FT METAL 55 55 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 SQ SEQUENCE 1342 AA; 147422 MW; 169254E4AFAAE021 CRC64;  
  
 Query Match 83.7%; Score 36; DB 1; Length 1342;  
 Best Local Similarity 71.4%; Pred. No. 30;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
 Qy 1 CGVRLGC 7  
 || :|||  
 Db 41 CGTKLGC 47

# RESULT 7

XDH\_DROSU  
 ID XDH\_DROSU STANDARD; PRT; 1344 AA.  
 AC P91711;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Xanthine dehydrogenase (EC 1.1.1.204) (XD) (Rosy locus protein).  
 GN RY OR XDH.  
 OS Drosophila subobscura (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7241;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97070823; PubMed=8913749;  
 RA Comeron J.M., Aguade M.;  
 RT "Synonymous substitutions in the Xdh gene of Drosophila:  
 RT heterogeneous distribution along the coding region.";  
 RL Genetics 144:1053-1062(1996).  
 CC -!- CATALYTIC ACTIVITY: Xanthine + NAD(+) + H(2)O = urate + NADH.  
 CC -!- COFACTOR: FAD, MOLYBDOPTERIN, AND TWO 2FE-2S CLUSTERS.  
 CC -!- SUBUNIT: Homodimer (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Peroxisomal.  
 CC -!- SIMILARITY: BELONGS TO THE XANTHINE DEHYDROGENASE FAMILY.  
 CC -!- SIMILARITY: TO 2FE-2S FERREDOXINS IN THE N-TERMINAL DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Y08237; CAA69405.1; -.

DR HSSP; P80457; 1FO4.  
 DR FlyBase; FBgn0013892; Dsub\ry.  
 DR InterPro; IPR002888; 2Fe-2S\_bind.  
 DR InterPro; IPR006058; 2Fe2S\_ferredoxin.  
 DR InterPro; IPR000674; Aldxan\_dh\_C.  
 DR InterPro; IPR005107; CO\_deh\_flav\_C.  
 DR InterPro; IPR002346; dehydrog\_molyb.  
 DR InterPro; IPR000572; Euk\_Mb\_oxred.  
 DR InterPro; IPR001041; Ferredoxin.  
 DR Pfam; PF02738; Ald\_Xan\_dh\_C2; 1.  
 DR Pfam; PF01315; Ald\_Xan\_dh\_C; 1.  
 DR Pfam; PF03450; CO\_deh\_flav\_C; 1.  
 DR Pfam; PF00941; FAD\_binding\_5; 1.  
 DR Pfam; PF00111; fer2; 1.  
 DR Pfam; PF01799; fer2\_2; 1.  
 DR ProDom; PD186071; 2Fe-2S\_bind; 1.  
 DR PROSITE; PS00197; 2FE2S\_FERREDOXIN; 1.  
 DR PROSITE; PS00559; MOLYBDOPTERIN\_EUK; 1.  
 KW Oxidoreductase; NAD; Molybdenum; Flavoprotein; FAD; Metal-binding;  
 KW Iron-sulfur; Iron; 2Fe-2S; Peroxisome.  
 FT METAL 42 42 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 FT METAL 48 48 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 FT METAL 53 53 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 FT METAL 56 56 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 SQ SEQUENCE 1344 AA; 147254 MW; 1DDB5BAC0E4C3175 CRC64;  
  
 Query Match 83.7%; Score 36; DB 1; Length 1344;  
 Best Local Similarity 71.4%; Pred. No. 30;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
 Qy 1 CGVRLGC 7  
 || :|||  
 Db 42 CGTKLGC 48

# RESULT 8

XDH\_CALVI  
 ID XDH\_CALVI STANDARD; PRT; 1353 AA.  
 AC P08793;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Xanthine dehydrogenase (EC 1.1.1.204) (XD).  
 GN XDH.  
 OS Calliphora vicina (Blue blowfly) (Calliphora erythrocephala).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Oestroidea;  
 OC Calliphoridae; Calliphora.  
 OX NCBI\_TaxID=7373;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90185213; PubMed=2516831;  
 RA Houde M., Tiveron M.C., Bregegere F.;  
 RT "Divergence of the nucleotide sequences encoding xanthine  
 RT dehydrogenase in Calliphora vicina and Drosophila melanogaster."  
 RL Gene 85:391-402(1989).  
 RN [2]

RP SEQUENCE OF 208-367 FROM N.A.  
 RX MEDLINE=88137956; PubMed=2830167;  
 RA Rocher-Chambonnet C., Berreur P., Houde M., Tiveron M.C.,  
 RA Lepesant J.-A., Bregegere F.;  
 RT "Cloning and partial characterization of the xanthine dehydrogenase  
 RT gene of *Calliphora vicina*, a distant relative of *Drosophila*  
 RT *melanogaster*.";   
 RL Gene 59:201-212(1987).  
 CC -!- CATALYTIC ACTIVITY: Xanthine + NAD(+) + H(2)O = urate + NADH.  
 CC -!- COFACTOR: FAD, MOLYBDOPTERIN, AND TWO 2FE-2S CLUSTERS.  
 CC -!- SUBUNIT: Homodimer.  
 CC -!- SUBCELLULAR LOCATION: Peroxisomal.  
 CC -!- SIMILARITY: BELONGS TO THE XANTHINE DEHYDROGENASE FAMILY.  
 CC -!- SIMILARITY: TO 2FE-2S FERREDOXINS IN THE N-TERMINAL DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X07229; CAA30189.1; -.  
 DR EMBL; X07323; CAA30281.1; -.  
 DR EMBL; X07324; CAA30281.1; JOINED.  
 DR EMBL; X07325; CAA30281.1; JOINED.  
 DR EMBL; M18423; AAA27879.1; -.  
 DR PIR; JQ0407; JQ0407.  
 DR HSSP; P80457; 1FO4.  
 DR InterPro; IPR002888; 2Fe-2S\_bind.  
 DR InterPro; IPR006058; 2Fe2S\_ferredoxin.  
 DR InterPro; IPR000674; Aldxan\_dh\_C.  
 DR InterPro; IPR005107; CO\_deh\_flav\_C.  
 DR InterPro; IPR002346; dehydrog\_molyb.  
 DR InterPro; IPR000572; Euk\_Mb\_oxred.  
 DR InterPro; IPR001041; Ferredoxin.  
 DR Pfam; PF02738; Ald\_Xan\_dh\_C2; 1.  
 DR Pfam; PF01315; Ald\_Xan\_dh\_C; 1.  
 DR Pfam; PF03450; CO\_deh\_flav\_C; 1.  
 DR Pfam; PF00941; FAD\_binding\_5; 1.  
 DR Pfam; PF00111; fer2; 1.  
 DR Pfam; PF01799; fer2\_2; 1.  
 DR ProDom; PD186071; 2Fe-2S\_bind; 1.  
 DR PROSITE; PS00197; 2FE2S\_FERREDOXIN; 1.  
 DR PROSITE; PS00559; MOLYBDOPTERIN\_EUK; 1.  
 KW Oxidoreductase; NAD; Molybdenum; Flavoprotein; FAD; Metal-binding;  
 KW Iron-sulfur; Iron; 2Fe-2S; Peroxisome.  
 FT METAL 50 50 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 FT METAL 56 56 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 FT METAL 61 61 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 FT METAL 64 64 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 SQ SEQUENCE 1353 AA; 150208 MW; 7120361C57C3E297 CRC64;

Query Match 83.7%; Score 36; DB 1; Length 1353;  
 Best Local Similarity 71.4%; Pred. No. 30;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
 || :|||  
 Db 50 CGTKLGC 56

RESULT 9

XDH\_CHICK

ID XDH\_CHICK STANDARD; PRT; 1358 AA.  
 AC P47990;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Xanthine dehydrogenase/oxidase [Includes: Xanthine dehydrogenase  
 DE (EC 1.1.1.204) (XD); Xanthine oxidase (EC 1.1.3.22) (XO) (Xanthine  
 DE oxidoreductase)].  
 GN XDH.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=95155354; PubMed=7852355;  
 RA Satoh A., Amaya Y., Noda K., Nishino T.;  
 RT "The structure of chicken liver xanthine dehydrogenase. cDNA cloning  
 RT and the domain structure."  
 RL J. Biol. Chem. 270:2818-2826(1995).  
 CC -!- FUNCTION: THIS ENZYME CAN BE CONVERTED FROM THE DEHYDROGENASE FORM  
 CC (D) TO THE OXIDASE FORM (O) IRREVERSIBLY BY PROTEOLYSIS OR  
 CC REVERSIBLY THROUGH THE OXIDATION OF SULFHYDRYL GROUPS.  
 CC -!- CATALYTIC ACTIVITY: Xanthine + NAD(+) + H(2)O = urate + NADH.  
 CC -!- CATALYTIC ACTIVITY: Xanthine + H(2)O + O(2) = urate + H(2)O(2).  
 CC -!- COFACTOR: FAD, MOLYBDOPTERIN, AND TWO 2FE-2S CLUSTERS.  
 CC -!- SUBUNIT: Homodimer.  
 CC -!- SUBCELLULAR LOCATION: Peroxisomal.  
 CC -!- SIMILARITY: BELONGS TO THE XANTHINE DEHYDROGENASE FAMILY.  
 CC -!- SIMILARITY: TO 2FE-2S FERREDOXINS IN THE N-TERMINAL DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D13221; BAA02502.1; -.  
 DR PIR; A55711; XOCHDH.  
 DR HSSP; P80457; 1FO4.  
 DR InterPro; IPR002888; 2Fe-2S\_bind.  
 DR InterPro; IPR006058; 2Fe2S\_ferredoxin.  
 DR InterPro; IPR000674; Aldxan\_dh\_C.  
 DR InterPro; IPR005107; CO\_deh\_flav\_C.  
 DR InterPro; IPR002346; dehydrog\_molyb.

DR InterPro; IPR000572; Euk\_Mb\_oxred.  
 DR InterPro; IPR001041; Ferredoxin.  
 DR Pfam; PF02738; Ald\_Xan\_dh\_C2; 1.  
 DR Pfam; PF01315; Ald\_Xan\_dh\_C; 1.  
 DR Pfam; PF03450; CO\_deh\_flav\_C; 1.  
 DR Pfam; PF00941; FAD\_binding\_5; 1.  
 DR Pfam; PF00111; fer2; 1.  
 DR Pfam; PF01799; fer2\_2; 1.  
 DR ProDom; PD186071; 2Fe-2S\_bind; 1.  
 DR PROSITE; PS00197; 2FE2S\_FERREDOXIN; 1.  
 DR PROSITE; PS00559; MOLYBDOPTERIN\_EUK; 1.  
 KW Oxidoreductase; NAD; Molybdenum; Flavoprotein; FAD; Metal-binding;  
 KW Iron-sulfur; Iron; 2Fe-2S.  
 FT METAL 41 41 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 FT METAL 47 47 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 FT METAL 52 52 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 FT METAL 55 55 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 SQ SEQUENCE 1358 AA; 149613 MW; 53B049B38704995F CRC64;

Query Match 83.7%; Score 36; DB 1; Length 1358;  
 Best Local Similarity 71.4%; Pred. No. 30;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
 || :|||  
 Db 41 CGTKLGC 47

#### RESULT 10

##### NEU2\_STRCA

ID NEU2\_STRCA STANDARD; PRT; 132 AA.  
 AC P21916;  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 01-MAY-1991 (Rel. 18, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Neurophysin 2 (MSEL-neurophysin).  
 OS Struthio camelus (Ostrich).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Palaeognathae; Struthioniformes; Struthionidae;  
 OC Struthio.  
 OX NCBI\_TaxID=8801;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=89254272; PubMed=2722398;  
 RA Lazure C., Saayman H.S., Naude R.J., Oelofsen W., Chretien M.;  
 RT "Ostrich MSEL-neurophysin belongs to the class of two-domain 'big'  
 RT neurophysin as indicated by complete amino acid sequence of the  
 RT neurophysin/copeptin.";  
 RL Int. J. Pept. Protein Res. 33:46-58(1989).  
 CC -!- FUNCTION: Neurophysin 2 specifically binds vasopressin.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- MISCELLANEOUS: IN NON-MAMMALIAN TETRAPODS, THE PROTEOLYTIC  
 CC PROCESSING OF THE PRO-VASOTOCIN INVOLVES ONLY ONE CLEAVAGE,  
 CC RELEASING THE HORMONE MOIETY AND A "BIG" NEUROPHYSIN WITH TWO  
 CC DOMAINS HOMOLOGOUS TO THE MAMMALIAN NEUROPHYSIN II AND COPEPTIN,  
 CC RESPECTIVELY.  
 CC -!- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.



DR PIR; A30978; A30978.  
 DR HSSP; P01180; 1NPO.  
 DR InterPro; IPR000981; Neurhyp\_horm.  
 DR Pfam; PF00184; hormone5; 1.  
 DR PRINTS; PR00831; NEUROPHYSIN.  
 DR ProDom; PD001676; Neurhyp\_horm; 1.  
 DR SMART; SM00003; NH; 1.  
 KW Hypothalamus; Cleavage on pair of basic residues.  
 FT DISULFID 10 54 BY SIMILARITY.  
 FT DISULFID 13 27 BY SIMILARITY.  
 FT DISULFID 21 44 BY SIMILARITY.  
 FT DISULFID 28 34 BY SIMILARITY.  
 FT DISULFID 61 73 BY SIMILARITY.  
 FT DISULFID 67 85 BY SIMILARITY.  
 FT DISULFID 74 79 BY SIMILARITY.  
 SQ SEQUENCE 132 AA; 13363 MW; D1BAC646D58CB33E CRC64;

Query Match 79.1%; Score 34; DB 1; Length 132;  
 Best Local Similarity 71.4%; Pred. No. 9.2;  
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
 || |||  
 Db 28 CGAELGC 34

RESULT 11  
 NEUV\_CHICK

ID NEUV\_CHICK STANDARD; PRT; 161 AA.  
 AC P24787;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Vasotocin-neurophysin VT precursor.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=White leghorn; TISSUE=Hypothalamus;  
 RA Hunt N., Kluever D., Ivell R.;  
 RL Submitted (NOV-1990) to the EMBL/GenBank/DDBJ databases.  
 CC -!- FUNCTION: VASOTOCIN IS AN ANTIDIURETIC HORMONE.  
 CC -!- DOMAIN: IN NON-MAMMALIAN TETRAPODS, THE PROTEOLYTIC PROCESSING OF  
 CC THE PRO-VASOTOCIN INVOLVES ONLY ONE CLEAVAGE, RELEASING THE  
 CC HORMONE MOIETY AND A "BIG" NEUROPHYSIN WITH TWO DOMAINS HOMOLOGOUS  
 CC TO THE MAMMALIAN NEUROPHYSIN II AND COPEPTIN, RESPECTIVELY.  
 CC -!- PTM: SEVEN DISULFIDE BONDS ARE PRESENT IN NEUROPHYSIN.  
 CC -!- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X55130; CAA38923.1; -.  
 DR PIR; S14480; S14480.  
 DR HSSP; P01180; 1NPO.  
 DR InterPro; IPR000981; Neurhyp\_horm.  
 DR Pfam; PF00220; hormone4; 1.  
 DR Pfam; PF00184; hormone5; 1.  
 DR PRINTS; PR00831; NEUROPHYSIN.  
 DR ProDom; PD001676; Neurhyp\_horm; 1.  
 DR SMART; SM00003; NH; 1.  
 DR PROSITE; PS00264; NEUROHYPOPHYS\_HORM; 1.  
 KW Hormone; Hypothalamus; Cleavage on pair of basic residues;  
 KW Amidation; Signal.  
 FT SIGNAL 1 19  
 FT PEPTIDE 20 28 VASOTOCIN.  
 FT PEPTIDE 32 161 VT NEUROPHYSIN.  
 FT DISULFID 20 25 BY SIMILARITY.  
 FT MOD\_RES 28 28 AMIDATION (G-29 PROVIDE AMIDE GROUP).  
 SQ SEQUENCE 161 AA; 16693 MW; 2802FBBED5E52277 CRC64;

Query Match 79.1%; Score 34; DB 1; Length 161;  
 Best Local Similarity 71.4%; Pred. No. 11;  
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
 || |||  
 Db 59 CGAELGC 65

# RESULT 12

## BIOD\_XANAC

ID BIOD\_XANAC STANDARD; PRT; 224 AA.  
 AC Q8PGK0;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Dethiobiotin synthetase (EC 6.3.3.3) (Dethiobiotin synthase) (DTB  
 DE synthetase) (DTBS).  
 GN BIOD OR XAC3616.  
 OS Xanthomonas axonopodis (pv. citri).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xanthomonas.  
 OX NCBI\_TaxID=92829;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=306 / ATCC 13902 / XV 101;  
 RX MEDLINE=22022145; PubMed=12024217;  
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,  
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,  
 RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,  
 RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,  
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,  
 RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,

RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,  
 RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,  
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,  
 RA Setubal J.C., Kitajima J.P.;  
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing  
 RT host specificities.";  
 RL Nature 417:459-463(2002).  
 CC -!- CATALYTIC ACTIVITY: ATP + 7,8-diaminononanoate + CO(2) = ADP +  
 CC phosphate + dethiobiotin.  
 CC -!- COFACTOR: Magnesium (By similarity).  
 CC -!- PATHWAY: Bioconversion of pimelate into dethiobiotin.  
 CC -!- SIMILARITY: BELONGS TO THE DETHIOBIOTIN SYNTHETASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE012012; AAM38459.1; -.  
 DR HAMAP; MF\_00336; -; 1.  
 DR InterPro; IPR004472; BioD.  
 DR InterPro; IPR002586; CbiA\_P.  
 DR Pfam; PF01656; CbiA; 1.  
 DR TIGRFAMs; TIGR00347; bioD; 1.  
 KW Biotin biosynthesis; Ligase; Magnesium; ATP-binding;  
 KW Complete proteome.  
 FT NP\_BIND 10 18 ATP (BY SIMILARITY).  
 SQ SEQUENCE 224 AA; 23719 MW; D0FEF451A315C7CE CRC64;

Query Match 79.1%; Score 34; DB 1; Length 224;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GVRLGC 7  
 |||||  
 Db 148 GVRLGC 153

RESULT 13  
 BIOD\_XANCP  
 ID BIOD\_XANCP STANDARD; PRT; 224 AA.  
 AC Q8PCW4;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Dethiobiotin synthetase (EC 6.3.3.3) (Dethiobiotin synthase) (DTB  
 DE synthetase) (DTBS).  
 GN BIOD OR XCC0587.  
 OS Xanthomonas campestris (pv. campestris).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xanthomonas.

OX NCBI\_TaxID=340;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 33913 / NCPPB 528;  
 RX MEDLINE=22022145; PubMed=12024217;  
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,  
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,  
 RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,  
 RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,  
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,  
 RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
 RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,  
 RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,  
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,  
 RA Setubal J.C., Kitajima J.P.;  
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing  
 RT host specificities.";  
 RL Nature 417:459-463(2002).  
 CC -!- CATALYTIC ACTIVITY: ATP + 7,8-diaminononanoate + CO(2) = ADP +  
 CC phosphate + dethiobiotin.  
 CC -!- COFACTOR: Magnesium (By similarity).  
 CC -!- PATHWAY: Bioconversion of pimelate into dethiobiotin.  
 CC -!- SIMILARITY: BELONGS TO THE DETHIOBIOTIN SYNTHETASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE012156; AAM39903.1; -.  
 DR HAMAP; MF\_00336; -; 1.  
 DR InterPro; IPR004472; BioD.  
 DR InterPro; IPR002586; CbiA\_P.  
 DR Pfam; PF01656; CbiA; 1.  
 DR TIGRFAMs; TIGR00347; bioD; 1.  
 KW Biotin biosynthesis; Ligase; Magnesium; ATP-binding;  
 KW Complete proteome.  
 FT NP\_BIND 10 18 ATP (BY SIMILARITY).  
 SQ SEQUENCE 224 AA; 23651 MW; 903596EF3A032439 CRC64;

Query Match 79.1%; Score 34; DB 1; Length 224;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GVRLGC 7  
 |||||  
 Db 148 GVRLGC 153

RESULT 14

BIOD\_PSEAE

ID BIOD\_PSEAE STANDARD; PRT; 228 AA.  
AC Q9I614;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Dethiobiotin synthetase (EC 6.3.3.3) (Dethiobiotin synthase) (DTB  
DE synthetase) (DTBS).  
GN BIOD OR PA0504.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 15692 / PA01;  
RX MEDLINE=20437337; PubMed=10984043;  
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,  
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
RT opportunistic pathogen."  
RL Nature 406:959-964(2000).  
CC -!- CATALYTIC ACTIVITY: ATP + 7,8-diaminononanoate + CO(2) = ADP +  
CC phosphate + dethiobiotin.  
CC -!- COFACTOR: MAGNESIUM (BY SIMILARITY).  
CC -!- PATHWAY: Bioconversion of pimelate into dethiobiotin.  
CC -!- SIMILARITY: BELONGS TO THE DETHIOBIOTIN SYNTHETASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE004487; AAG03893.1; -.  
DR PIR; B83583; B83583.  
DR HSSP; P13000; 1DTS.  
DR HAMAP; MF\_00336; -; 1.  
DR InterPro; IPR004472; BioD.  
DR InterPro; IPR002586; CbiA\_P.  
DR Pfam; PF01656; CbiA; 1.  
DR TIGRFAMs; TIGR00347; bioD; 1.  
KW Biotin biosynthesis; Ligase; Magnesium; ATP-binding;  
KW Complete proteome.  
FT NP\_BIND 8 16 ATP (BY SIMILARITY).  
SQ SEQUENCE 228 AA; 23337 MW; 4CC964E353B3085A CRC64;

Query Match 79.1%; Score 34; DB 1; Length 228;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy            2 GVRLGC 7  
              |||||  
 Db           147 GVRLGC 152

RESULT 15

KVB3\_HUMAN

ID    KVB3\_HUMAN            STANDARD;            PRT;    404 AA.  
 AC    O43448;  
 DT    16-OCT-2001 (Rel. 40, Created)  
 DT    16-OCT-2001 (Rel. 40, Last sequence update)  
 DT    16-OCT-2001 (Rel. 40, Last annotation update)  
 DE    Voltage-gated potassium channel beta-3 subunit (K+ channel beta-3  
 DE    subunit) (Kv-beta-3).  
 GN    KCNAB3 OR KCNA3B.  
 OS    Homo sapiens (Human).  
 OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX    NCBI\_TaxID=9606;  
 RN    [1]  
 RP    SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.  
 RC    TISSUE=Brain;  
 RX    MEDLINE=99074289; PubMed=9857044;  
 RA    Leicher T., Baehring R., Isbrandt D., Pongs O.;  
 RT    "Coexpression of the KCNA3B gene product with Kv1.5 leads to a novel  
 RT    A-type potassium channel.";  
 RL    J. Biol. Chem. 273:35095-35101(1998).  
 CC    -!- FUNCTION: ACCESSORY POTASSIUM CHANNEL PROTEIN WHICH MODULATES THE  
 CC    ACTIVITY OF THE PORE-FORMING ALPHA SUBUNIT. ALTERS THE FUNCTIONAL  
 CC    PROPERTIES OF KV1.5.  
 CC    -!- SUBUNIT: FORMS HETEROMULTIMERIC COMPLEX WITH ALPHA SUBUNITS.  
 CC    -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
 CC    -!- TISSUE SPECIFICITY: BRAIN-SPECIFIC EXPRESSION. MOST PROMINENT  
 CC    EXPRESSION IN CEREBELLUM. WEAKER SIGNALS DETECTED IN CORTEX,  
 CC    OCCIPITAL LOBE, FRONTAL LOBE AND TEMPORAL LOBE. NOT DETECTED IN  
 CC    SPINAL CORD, HEART, LUNG, LIVER, KIDNEY, PANCREAS, PLACENTA AND  
 CC    SKELETAL MUSCLE.  
 CC    -!- DOMAIN: ALTERATION OF FUNCTIONAL PROPERTIES OF ALPHA SUBUNIT IS  
 CC    MEDIATED THROUGH N-TERMINAL DOMAIN OF BETA SUBUNIT (PROBABLE).  
 CC    -!- SIMILARITY: BELONGS TO THE SHAKER POTASSIUM CHANNEL BETA SUBUNIT  
 CC    FAMILY.  
 CC    -----  
 CC    This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC    the European Bioinformatics Institute. There are no restrictions on its  
 CC    use by non-profit institutions as long as its content is in no way  
 CC    modified and this statement is not removed. Usage by and for commercial  
 CC    entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC    or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC    -----  
 DR    EMBL; AF016411; AAB92499.1; -.  
 DR    Genew; HGNC:6230; KCNAB3.  
 DR    MIM; 604111; -.  
 DR    GO; GO:0015459; F:potassium channel regulator activity; TAS.  
 DR    GO; GO:0006813; P:potassium ion transport; TAS.  
 DR    InterPro; IPR001395; Aldo/ket\_red.

DR InterPro; IPR005402; KCNAB3\_channel.  
DR InterPro; IPR005399; KCNAB\_channel.  
DR InterPro; IPR005983; KCNAB\_core.  
DR Pfam; PF00248; aldo\_ket\_red; 1.  
DR PRINTS; PR01580; KCNAB3CHANNEL.  
DR PRINTS; PR01577; KCNABCHANNEL.  
DR ProDom; PD000288; Aldo/ket\_red; 2.  
DR TIGRFAMs; TIGR01293; Kv\_beta; 1.  
KW Ionic channel; Ion transport; Potassium transport;  
KW Voltage-gated channel.  
SQ SEQUENCE 404 AA; 43530 MW; 08265CC07929A1BA CRC64;

Query Match 79.1%; Score 34; DB 1; Length 404;  
Best Local Similarity 71.4%; Pred. No. 25;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
|||:|  
Db 86 CGVRVSC 92

Search completed: November 13, 2003, 09:46:35  
Job time : 5.01042 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:31:40 ; Search time 18.4479 Seconds  
(without alignments)  
97.917 Million cell updates/sec

Title: US-09-228-866-6  
Perfect score: 43  
Sequence: 1 CGVRLGC 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*

```

5:  sp_invertebrate:*
6:  sp_mammal:*
7:  sp_mhc:*
8:  sp_organelle:*
9:  sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	38	88.4	248	2	Q8GPS3	Q8gps3 pseudomonas
2	37	86.0	446	3	Q96VM9	Q96vm9 fusarium ox
3	36	83.7	397	16	Q9X224	Q9x224 thermotoga
4	36	83.7	438	5	Q9VUX5	Q9vux5 drosophila
5	36	83.7	552	4	Q9NPZ7	Q9npz7 homo sapien
6	36	83.7	662	6	Q9BDF6	Q9bdf6 equus cabal
7	36	83.7	678	11	Q8VDT1	Q8vdt1 mus musculu
8	36	83.7	685	11	Q8BZW1	Q8bzw1 mus musculu
9	36	83.7	685	11	Q8BGU9	Q8bgu9 mus musculu
10	36	83.7	1326	5	Q23829	Q23829 calliphora
11	36	83.7	1335	5	Q9VFZ9	Q9vzf9 drosophila
12	36	83.7	1347	5	Q9BIF9	Q9bif9 ceratitidis c
13	36	83.7	1936	5	Q9VWJ6	Q9vwj6 drosophila
14	36	83.7	2270	12	Q9JFN3	Q9jfn3 tupaia para
15	35	81.4	322	5	Q9VAN9	Q9van9 drosophila
16	35	81.4	342	2	Q9WXG6	Q9wxg6 alcaligenes
17	35	81.4	575	10	Q9AUY1	Q9auy1 oryza sativ
18	35	81.4	903	4	Q8TDY4	Q8tdy4 homo sapien
19	35	81.4	3853	5	Q8IJW2	Q8ijw2 plasmodium
20	34	79.1	148	2	Q8KTK2	Q8ktk2 sinorhizobi
21	34	79.1	167	10	Q8S5N1	Q8s5n1 oryza sativ
22	34	79.1	173	8	Q8M0F2	Q8m0f2 phoxinus eo
23	34	79.1	267	16	Q9PDN8	Q9pdn8 xylella fas
24	34	79.1	317	13	Q9DGR3	Q9dgr3 xenopus lae
25	34	79.1	382	10	O82594	O82594 arabidopsis
26	34	79.1	388	10	Q8SAW1	Q8saw1 oryza sativ
27	34	79.1	421	16	Q92SY4	Q92sy4 rhizobium m
28	34	79.1	540	5	Q9VCL3	Q9vcl3 drosophila
29	34	79.1	706	10	Q8S5J1	Q8s5j1 oryza sativ
30	34	79.1	772	5	O60958	O60958 leishmania
31	34	79.1	791	2	Q9L5R1	Q9l5r1 salmonella
32	34	79.1	809	16	Q935R1	Q935r1 salmonella
33	33	76.7	80	8	O47957	O47957 phoxinus eo
34	33	76.7	173	8	Q8WB67	Q8wb67 phoxinus er



35	33	76.7	173	8	O79951	O79951 osmerus mor
36	33	76.7	173	8	Q9BA03	Q9ba03 gonostoma g
37	33	76.7	173	8	Q8LUL1	Q8lul1 phoxinus er
38	33	76.7	173	8	Q951J1	Q951j1 phoxinus er
39	33	76.7	173	8	Q951I9	Q951i9 phoxinus er
40	33	76.7	173	8	Q8WB69	Q8wb69 phoxinus er
41	33	76.7	174	8	O78792	O78792 osmerus mor
42	33	76.7	184	6	Q95JG8	Q95jg8 bos taurus
43	33	76.7	200	11	Q64657	Q64657 rattus sp.
44	33	76.7	262	2	O85961	O85961 sphingomona
45	33	76.7	312	6	Q9TTR6	Q9ttr6 bos taurus

# ALIGNMENTS

## RESULT 1

Q8GPS3

ID Q8GPS3 PRELIMINARY; PRT; 248 AA.  
AC Q8GPS3;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Conserved hypothetical protein.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SG17M;  
RX MEDLINE=22313472; PubMed=12426355;  
RA Larbig K.D., Christmann A., Johann A., Klockgether J., Hartsch T.,  
RA Merkl R., Wiehlmann L., Fritz H.J., Tummler B.;  
RT "Gene Islands Integrated into tRNA(Gly) Genes Confer Genome Diversity  
RT on a Pseudomonas aeruginosa Clone.";  
RL J. Bacteriol. 184:6665-6680(2002).  
DR EMBL; AF440524; AAN62301.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 248 AA; 27423 MW; CD2495170A05109C CRC64;

Query Match 88.4%; Score 38; DB 2; Length 248;  
Best Local Similarity 85.7%; Pred. No. 15;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	CGVRLGC	7
Db	23	CGVRAGC	29

## RESULT 2

Q96VM9

ID Q96VM9 PRELIMINARY; PRT; 446 AA.  
AC Q96VM9;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Putative transposase.  
 OS Fusarium oxysporum f. sp. ciceris.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Hypocreales; mitosporic Hypocreales; Fusarium.  
 OX NCBI\_TaxID=62683;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=8012; TRANSPOSON=Fotci;  
 RA Horman S.R., Bainbridge B.W.;  
 RT "Fotci, a hAT family transposable element in Fusarium oxysporum f. sp.  
 RT ciceris.";  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AY039810; AAK82929.1; -.  
 DR InterPro; IPR004875; CENP-B.  
 DR InterPro; IPR006600; CENPB.  
 DR Pfam; PF03184; DDE; 1.  
 DR SMART; SM00674; CENPB; 1.  
 SQ SEQUENCE 446 AA; 50489 MW; B5F1862F7F01ED8A CRC64;

Query Match 86.0%; Score 37; DB 3; Length 446;  
 Best Local Similarity 85.7%; Pred. No. 39;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
 |||| ||  
 Db 430 CGVRQGC 436

# RESULT 3

Q9X224

ID Q9X224 PRELIMINARY; PRT; 397 AA.  
 AC Q9X224;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Aspartate aminotransferase.  
 GN TM1698.  
 OS Thermotoga maritima.  
 OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.  
 OX NCBI\_TaxID=2336;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MSB8 / DSM 3109;  
 RX MEDLINE=99287316; PubMed=10360571;  
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,  
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,  
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;  
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from  
 RT genome sequence of Thermotoga maritima.";  
 RL Nature 399:323-329(1999).  
 DR EMBL; AE001810; AAD36765.1; -.  
 DR TIGR; TM1698; -.  
 DR InterPro; IPR001176; ACC\_synthase.  
 DR InterPro; IPR004839; Aminotransf1/2.

DR InterPro; IPR004838; NHtransf\_1.  
 DR Pfam; PF00155; aminotran\_1\_2; 1.  
 DR PRINTS; PR00753; ACCSYNTHASE.  
 DR PROSITE; PS00105; AA\_TRANSFER\_CLASS\_1; 1.  
 KW Transferase; Aminotransferase; Complete proteome.  
 SQ SEQUENCE 397 AA; 44917 MW; F49307EE068223DC CRC64;

Query Match 83.7%; Score 36; DB 16; Length 397;  
 Best Local Similarity 71.4%; Pred. No. 54;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
 || |:||  
 Db 237 CGARVGC 243

#### RESULT 4

Q9VUX5

ID Q9VUX5 PRELIMINARY; PRT; 438 AA.  
 AC Q9VUX5;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Th gene product.  
 GN TH OR CG12284.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*."  
 RL Science 287:2185-2195(2000).  
 CC -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.  
 DR EMBL; AE003528; AAG22319.1; -.  
 DR HSSP; Q13490; 1QBH.  
 DR FlyBase; FBgn0003691; th.  
 DR InterPro; IPR001370; BIR.  
 DR InterPro; IPR001841; Znf\_ring.  
 DR Pfam; PF00653; BIR; 2.  
 DR Pfam; PF00097; zf-C3HC4; 1.  
 DR SMART; SM00238; BIR; 2.  
 DR SMART; SM00184; RING; 1.  
 DR PROSITE; PS01282; BIR\_REPEAT\_1; 2.  
 DR PROSITE; PS50143; BIR\_REPEAT\_2; 2.  
 DR PROSITE; PS50089; ZF\_RING\_2; 1.  
 KW Metal-binding; Zinc; Zinc-finger.  
 SQ SEQUENCE 438 AA; 48038 MW; 24CA8BC13F5DEF31 CRC64;

Query Match 83.7%; Score 36; DB 5; Length 438;  
 Best Local Similarity 71.4%; Pred. No. 60;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
 ||| :||  
 Db 83 CGVEIGC 89

# RESULT 5

Q9NPZ7

ID Q9NPZ7 PRELIMINARY; PRT; 552 AA.  
 AC Q9NPZ7;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE DJ1024N4.1 (Novel sodium:solute symporter family member similar to  
 DE SLC5A1 (SGLT1)) (Fragment).  
 GN DJ1024N4.1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Coville G.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AL109659; CAC00574.1; -.  
 DR InterPro; IPR001734; Na/solut\_sympport.  
 DR Pfam; PF00474; SSF; 1.  
 DR TIGRFAMs; TIGR00813; sss; 1.  
 DR PROSITE; PS00456; NA\_SOLUT\_SYMP\_1; 1.  
 DR PROSITE; PS50283; NA\_SOLUT\_SYMP\_3; 1.  
 FT NON\_TER 552 552  
 SQ SEQUENCE 552 AA; 59853 MW; C5D88CA8548EAEA7 CRC64;

Query Match 83.7%; Score 36; DB 4; Length 552;  
 Best Local Similarity 71.4%; Pred. No. 73;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
 |||:|  
 Db 352 CGARVGC 358

#### RESULT 6

##### Q9BDF6

ID Q9BDF6 PRELIMINARY; PRT; 662 AA.  
 AC Q9BDF6;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Na+/glucose co-transporter.  
 GN SGLT1.  
 OS Equus caballus (Horse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
 OX NCBI\_TaxID=9796;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Dyer J.;  
 RT "Molecular characterisation of carbohydrate digestion and absorption  
 RT in equine small intestine."  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AJ292081; CAC35538.1; -.  
 DR InterPro; IPR001734; Na/solut\_sympport.  
 DR Pfam; PF00474; SSF; 1.  
 DR TIGRFAMs; TIGR00813; sss; 1.  
 DR PROSITE; PS00456; NA\_SOLUT\_SYMP\_1; 1.  
 DR PROSITE; PS00457; NA\_SOLUT\_SYMP\_2; 1.  
 DR PROSITE; PS50283; NA\_SOLUT\_SYMP\_3; 1.  
 SQ SEQUENCE 662 AA; 73050 MW; 273C503BCD057631 CRC64;

Query Match 83.7%; Score 36; DB 6; Length 662;  
 Best Local Similarity 57.1%; Pred. No. 86;  
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
 ||::||  
 Db 355 CGIKVG 361

#### RESULT 7

## Q8VDT1

ID Q8VDT1 PRELIMINARY; PRT; 678 AA.  
AC Q8VDT1;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Similar to solute carrier family 5 (Sodium/glucose cotransporter),  
DE member 1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Strausberg R.;  
RL Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; BC021357; AAH21357.1; -.  
DR InterPro; IPR001734; Na/solut\_sympoort.  
DR Pfam; PF00474; SSF; 1.  
DR TIGRFAMs; TIGR00813; sss; 1.  
DR PROSITE; PS00456; NA\_SOLUT\_SYMP\_1; 1.  
DR PROSITE; PS50283; NA\_SOLUT\_SYMP\_3; 1.  
SQ SEQUENCE 678 AA; 74250 MW; CB871C9F182A626D CRC64;

Query Match 83.7%; Score 36; DB 11; Length 678;  
Best Local Similarity 71.4%; Pred. No. 88;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
|| |:||  
Db 348 CGARVGC 354

## RESULT 8

## Q8BZW1

ID Q8BZW1 PRELIMINARY; PRT; 685 AA.  
AC Q8BZW1;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Weakly similar to NA+-glucose cotransporter type 1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Colon;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.";  
RL Nature 420:563-573(2002).  
DR EMBL; AK033425; BAC28283.1; -.  
SQ SEQUENCE 685 AA; 75035 MW; 7C332F95988BFC5C CRC64;

Query Match 83.7%; Score 36; DB 11; Length 685;  
Best Local Similarity 71.4%; Pred. No. 89;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
|||:  
Db 355 CGARVGC 361

RESULT 9

Q8BGU9

ID Q8BGU9 PRELIMINARY; PRT; 685 AA.  
AC Q8BGU9;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Weakly similar to NA+-glucose cotransporter type 1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Body, and Skin;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs."  
RL Nature 420:563-573(2002).  
DR EMBL; AK029158; BAC26332.1; -.  
DR EMBL; AK050696; BAC34386.1; -.  
SQ SEQUENCE 685 AA; 75065 MW; 6D223B4EE896572C CRC64;

Query Match 83.7%; Score 36; DB 11; Length 685;  
Best Local Similarity 71.4%; Pred. No. 89;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
|||:  
Db 355 CGARVGC 361

RESULT 10

Q23829

ID Q23829 PRELIMINARY; PRT; 1326 AA.  
AC Q23829;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Xanthine dehydrogenase (Xdh) gene allele 1, exons 2-4 (Fragment).  
OS Calliphora vicina (Blue blowfly) (Calliphora erythrocephala).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Oestroidea;  
OC Calliphoridae; Calliphora.  
OX NCBI\_TaxID=7373;  
RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=90185213; PubMed=2516831;  
 RA Houde M., Tiveron M.-C., Bregegere F.;  
 RT "Divergence of the nucleotide sequences encoding xanthine  
 RT dehydrogenase in *Calliphora vicina* and *Drosophila melanogaster*.";   
 RL Gene 85:391-402(1989).  
 CC -!- COFACTOR: BINDS 1 2FE-2S CLUSTER (BY SIMILARITY).  
 DR EMBL; M30316; AAA27880.1; -.  
 DR HSSP; P80457; 1FO4.  
 DR InterPro; IPR002888; 2Fe-2S\_bind.  
 DR InterPro; IPR006058; 2Fe2S\_ferredoxin.  
 DR InterPro; IPR000674; Aldxan\_dh\_C.  
 DR InterPro; IPR005107; CO\_deh\_flav\_C.  
 DR InterPro; IPR002346; dehydrog\_molyb.  
 DR InterPro; IPR000572; Euk\_Mb\_oxred.  
 DR InterPro; IPR001041; Ferredoxin.  
 DR Pfam; PF01315; Ald\_Xan\_dh\_C; 1.  
 DR Pfam; PF02738; Ald\_Xan\_dh\_C2; 1.  
 DR Pfam; PF03450; CO\_deh\_flav\_C; 1.  
 DR Pfam; PF00941; FAD\_binding\_5; 1.  
 DR Pfam; PF00111; fer2; 1.  
 DR Pfam; PF01799; fer2\_2; 1.  
 DR ProDom; PD186071; 2Fe-2S\_bind; 1.  
 DR PROSITE; PS00197; 2FE2S\_FERREDOXIN; 1.  
 DR PROSITE; PS00559; MOLYBDOPTERIN\_EUK; 1.  
 KW Iron; Iron-sulfur.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 1326 AA; 147155 MW; F84B266DBE93CC4E CRC64;  
  
 Query Match 83.7%; Score 36; DB 5; Length 1326;  
 Best Local Similarity 71.4%; Pred. No. 1.6e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
 || :|||  
 Db 23 CGTKLGC 29

RESULT 11  
 Q9VFZ9  
 ID Q9VFZ9 PRELIMINARY; PRT; 1335 AA.  
 AC Q9VFZ9;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE RY gene product.  
 GN RY OR CG7642.  
 OS *Drosophila melanogaster* (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; *Drosophila*.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,



RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 CC -!- COFACTOR: BINDS 1 2FE-2S CLUSTER (BY SIMILARITY).  
 DR EMBL; AE003698; AAF54895.1; -.  
 DR HSSP; P80457; 1FO4.  
 DR FlyBase; FBgn0003308; ry.  
 DR InterPro; IPR002888; 2Fe-2S\_bind.  
 DR InterPro; IPR006058; 2Fe2S\_ferredoxin.  
 DR InterPro; IPR000674; Aldxan\_dh\_C.  
 DR InterPro; IPR005107; CO\_deh\_flav\_C.  
 DR InterPro; IPR002346; dehydrog\_molyb.  
 DR InterPro; IPR000572; Euk\_Mb\_oxred.  
 DR InterPro; IPR001041; Ferredoxin.  
 DR Pfam; PF01315; Ald\_Xan\_dh\_C; 1.  
 DR Pfam; PF02738; Ald\_Xan\_dh\_C2; 1.  
 DR Pfam; PF03450; CO\_deh\_flav\_C; 1.  
 DR Pfam; PF00941; FAD\_binding\_5; 1.  
 DR Pfam; PF00111; fer2; 1.  
 DR Pfam; PF01799; fer2\_2; 1.  
 DR ProDom; PD186071; 2Fe-2S\_bind; 1.  
 DR PROSITE; PS00197; 2FE2S\_FERREDOXIN; 1.  
 DR PROSITE; PS00559; MOLYBDOPTERIN\_EUK; 1.

KW Iron; Iron-sulfur.

SQ SEQUENCE 1335 AA; 146926 MW; 230368EA59B30AD8 CRC64;

Query Match 83.7%; Score 36; DB 5; Length 1335;

Best Local Similarity 71.4%; Pred. No. 1.6e+02;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7

|| :|||

Db 37 CGTKLGC 43

## RESULT 12

### Q9BIF9

ID Q9BIF9 PRELIMINARY; PRT; 1347 AA.

AC Q9BIF9;

DT 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

DE Xanthine dehydrogenase.

GN XDH.

OS *Ceratitis capitata* (Mediterranean fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Tephritoidea; Tephritidae; Ceratitis.

OX NCBI\_TaxID=7213;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Benakeion;

RX MEDLINE=21405530; PubMed=11514452;

RA Pitts R.J., Zwiebel L.J.;

RT "Isolation and Characterization of the Xanthine Dehydrogenase Gene of  
the Mediterranean Fruit Fly, *Ceratitis capitata*.";

RL Genetics 158:1645-1655(2001).

CC -!- COFACTOR: BINDS 1 2FE-2S CLUSTER (BY SIMILARITY).

DR EMBL; AY014961; AAG47345.1; -.

DR HSSP; P80457; 1FO4.

DR InterPro; IPR002888; 2Fe-2S\_bind.

DR InterPro; IPR006058; 2Fe2S\_ferredoxin.

DR InterPro; IPR000674; Aldxan\_dh\_C.

DR InterPro; IPR005107; CO\_deh\_flav\_C.

DR InterPro; IPR002346; dehydrog\_molyb.

DR InterPro; IPR000572; Euk\_Mb\_oxred.

DR InterPro; IPR001041; Ferredoxin.

DR InterPro; IPR001497; Methyltransf\_1.

DR Pfam; PF01315; Ald\_Xan\_dh\_C; 1.

DR Pfam; PF02738; Ald\_Xan\_dh\_C2; 1.

DR Pfam; PF03450; CO\_deh\_flav\_C; 1.

DR Pfam; PF00941; FAD\_binding\_5; 1.

DR Pfam; PF00111; fer2; 1.

DR Pfam; PF01799; fer2\_2; 1.

DR ProDom; PD186071; 2Fe-2S\_bind; 1.

DR PROSITE; PS00197; 2FE2S\_FERREDOXIN; 1.

DR PROSITE; PS00374; MGMT; 1.

DR PROSITE; PS00559; MOLYBDOPTERIN\_EUK; 1.

KW Iron; Iron-sulfur.

SQ SEQUENCE 1347 AA; 149145 MW; 44FDC59F38DAEDE0 CRC64;

Query Match 83.7%; Score 36; DB 5; Length 1347;  
Best Local Similarity 71.4%; Pred. No. 1.6e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
|| :|||  
Db 49 CGTKLGC 55

RESULT 13

Q9VWJ6

ID Q9VWJ6 PRELIMINARY; PRT; 1936 AA.  
AC Q9VWJ6;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE CG8002 protein.  
GN CG8002.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkeley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
 RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of *Drosophila melanogaster* genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RT "Annotation of *Drosophila melanogaster* genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AE003511; AAF48942.2; -.  
 DR FlyBase; FBgn0031006; CG8002.  
 SQ SEQUENCE 1936 AA; 214502 MW; 49789D604233D1DA CRC64;

Query Match 83.7%; Score 36; DB 5; Length 1936;  
 Best Local Similarity 85.7%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGVRLGC 7  
 |||||  
 Db 249 CGVRLDC 255

RESULT 14

Q9JFN3 ID Q9JFN3 PRELIMINARY; PRT; 2270 AA.

AC Q9JFN3; 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE RNA polymerase.

GN L.

OS Tupaiá paramyxovirus (TPMV).

OC Viruses; ssRNA negative-strand viruses; Mononegavirales;

OC Paramyxoviridae; Paramyxovirinae.

OX NCBI\_TaxID=92129;

RN [1]

RP SEQUENCE FROM N.A.

RA Tidona C.A., Darat G.;

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF079780; AAF63393.1; -.

DR InterPro; IPR002877; FtsJ.

DR InterPro; IPR007098; RNA\_pol\_monon.

DR InterPro; IPR001016; Viral\_RNA\_pol\_L.

DR Pfam; PF01728; FtsJ; 1.

DR Pfam; PF00946; Paramyx\_RNA\_pol; 1.

DR PROSITE; PS50524; RDRP\_MONONEGAVIRALTES; 1.

SQ SEQUENCE 2270 AA; 259394 MW; 037C728466E95D95 CRC64;

Query Match 83.7%; Score 36; DB 12; Length 2270;

Best Local Similarity 71.4%; Pred. No. 2.6e+02;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGVRLGC 7 ||:|

DB 452 CGMRFGC 458

RESULT 15

Q9VAN9 ID Q9VAN9 PRELIMINARY; PRT; 322 AA.

AC Q9VAN9; 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)

DE CG14511 protein.

GN CG14511.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Aamati P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.;

RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA	Abritl J.F., Agbayan A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu T., Beasley E.M.,
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA	Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA	Merkulov G., Miltshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA	Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA	Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA	Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA	Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RT	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT	"The genome sequence of <i>Drosophila melanogaster</i> ."
RL	Science 287:2185-2195(2000).
DR	EMBL; AE003768; AAF56864.1; -.
DR	FLYBase; FBgn0039641; CG14511.
SO	SEQUENCE 322 AA; 36056 MW; 94F48488E19A1F29 CRC64;

Query Match 81.4%; Score 35; DB 5; Length 322;  
 Best Local Similarity 85.7%; Pred. No. 70;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGVRLGC 7  
 DB 12 CGVFLGC 18

Search completed: November 13, 2003, 09:51:04  
 Job time : 21.4479 secs

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:39:50 ; Search time 8.3125 Seconds  
(without alignments)  
35.630 Million cell updates/sec

Title: US-09-228-866-6

Perfect score: 43

Sequence: 1 CGVRLGC 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*

1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43	100.0	7	1	US-08-526-710-6
2	43	100.0	7	3	US-08-862-855-6
3	43	100.0	7	3	US-09-226-985-6
4	43	100.0	7	4	US-09-227-906-6
5	36	83.7	662	1	US-07-841-651-4
6	34	79.1	15	4	US-09-403-089A-8
7	34	79.1	219	4	US-09-199-637A-415
8	34	79.1	229	4	US-09-252-991A-19859
9	33	76.7	35	3	US-08-851-843A-178
10	33	76.7	35	3	US-08-974-549A-297
11	33	76.7	35	3	US-08-854-050-178
Sequence 178, App					
Sequence 297, App					
Sequence 178, App					
Sequence 19859, A					
Sequence 415, App					
Sequence 8, Appli					
Sequence 4, Appli					
Sequence 6, Appli					
Sequence 6, Appli					
Sequence 6, Appli					
Sequence 6, Appli					

RESULT 1  
US-08-526-710-6  
Sequence 6, Application US/08526710  
Patent No. 5622699  
GENERAL INFORMATION:  
APPLICANT: Ruostalahti, Erkki  
APPLICANT: Pasqualini, Renata  
TITLE OF INVENTION: Method of Identifying Molecules That  
TITLE OF INVENTION: Home to a Selected Organ In Vivo  
NUMBER OF SEQUENCES: 44  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell and Flores  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: United States  
ZIP: 92122  
COMPUTER READABLE FORM:

## ALIGNMENTS

Sequence 178, App	US-09-430-323-178	4	35	76.7	33	12
Sequence 6877, Ap	US-09-107-532A-6877	4	327	76.7	33	13
Sequence 42, Appl	US-08-861-774E-42	3	212	74.4	32	14
Sequence 367, App	US-09-198-452A-367	4	336	74.4	32	15
Sequence 1073, Ap	US-09-198-452A-1073	4	568	74.4	32	16
Sequence 9, Appl	US-09-165-827C-9	4	70	72.1	31	17
Sequence 10, Appl	US-09-165-827C-10	4	70	72.1	31	18
Sequence 275, App	US-08-905-223-275	3	89	72.1	31	19
Sequence 10, Appl	US-09-249-542-10	4	112	72.1	31	20
Sequence 25944, A	US-09-252-991A-25944	4	161	72.1	31	21
Sequence 6, Appl	US-08-401-068-6	2	219	72.1	31	22
Sequence 6, Appl	US-08-846-338-6	2	219	72.1	31	23
Sequence 8, Appl	US-08-411-768B-8	3	236	72.1	31	24
Sequence 14, Appl	US-09-165-827C-14	4	284	72.1	31	25
Sequence 4, Appl	US-08-846-762-4	2	316	72.1	31	26
Sequence 2, Appl	US-09-165-827C-2	4	342	72.1	31	27
Sequence 28724, A	US-09-252-991A-28724	4	359	72.1	31	28
Sequence 18143, A	US-09-252-991A-18143	4	384	72.1	31	29
Sequence 17629, A	US-09-252-991A-17629	4	445	72.1	31	30
Sequence 26597, A	US-09-252-991A-26597	4	638	72.1	31	31
Sequence 2, Appl	US-07-841-651-2	1	672	72.1	31	32
Sequence 3, Appl	US-07-841-651-3	1	672	72.1	31	33
Sequence 10, Appl	US-09-370-368-10	3	701	72.1	31	34
Sequence 972, App	US-08-159-339A-972	3	9	69.8	30	35
Sequence 988, App	US-08-159-339A-988	3	10	69.8	30	36
Sequence 5629, App	US-09-134-001C-5629	4	81	69.8	30	37
Sequence 736, App	US-09-205-258-736	4	98	69.8	30	38
Sequence 394, App	US-08-311-731A-394	4	135	69.8	30	39
Sequence 357, App	US-09-199-637A-357	4	136	69.8	30	40
Sequence 17872, A	US-09-252-991A-17872	4	137	69.8	30	41
Sequence 9, Appl	US-09-124-141-9	3	203	69.8	30	42
Sequence 1020, Ap	US-09-205-258-1020	4	214	69.8	30	43
Sequence 17, Appl	US-09-124-141-17	3	250	69.8	30	44
Sequence 10, Appl	US-08-952-736A-10	4	309	69.8	30	45



RESULT 2  
 US-08-862-855-6  
 Sequence 6, Application US/08862855  
 Patent No. 6068829  
 GENERAL INFORMATION:  
 APPLICANT: Ruostahhti, Erkki  
 APPLICANT: Pasqualini, Renata  
 TITLE OF INVENTION: Method of Identifying Molecules That  
 TITLE OF INVENTION: Home to a Selected Organ In Vivo  
 NUMBER OF SEQUENCES: 44  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: Campbell & Flores LLP  
 STREET: 4370 La Jolla Village Drive, Suite 700  
 CITY: San Diego  
 STATE: California  
 COUNTRY: United States  
 ZIP: 92122  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent in Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/862,855  
 FILING DATE:

QY 1 CGVRLGC 7  
 DB 1 CGVRLGC 7  
 Query Match 100.0%; Score 43; DB 1; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 2.5e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

US-08-526-710-6  
 MOLECULE TYPE: peptide  
 TOPOLOGY: linear  
 TYPE: amino acid  
 LENGTH: 7 amino acids  
 SEQUENCE CHARACTERISTICS:  
 INFORMATION FOR SEQ ID NO: 6:  
 TELEFAX: (619) 535-8949  
 TELEPHONE: (619) 535-9001  
 TELECOMMUNICATION INFORMATION:  
 REFERENCE/DOCKET NUMBER: P-LJ 1779  
 REGISTRATION NUMBER: 31,815  
 NAME: Campbell, Cathryn A.  
 ATTORNEY/AGENT INFORMATION:  
 CLASSIFICATION: 435  
 FILING DATE: 11-SEP-1995  
 APPLICATION NUMBER: US/08/526,710  
 CURRENT APPLICATION DATA:  
 SOFTWARE: Patent in Release #1.0, Version #1.25  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 COMPUTER: IBM PC compatible  
 MEDIUM TYPE: Floppy disk

CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/526,710  
FILING DATE: 11-SEP-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/813,273  
FILING DATE: 10-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-LJ 2621  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-862-855-6

Query Match 100.0%; Score 43; DB 3; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2.5e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGVRLGC 7  
| | | | |  
DB 1 CGVRLGC 7

RESULT 3  
US-09-226-985-6  
Sequence 6, Application US/09226985  
Patent No. 6296832  
GENERAL INFORMATION:  
APPLICANT: Ruoslahti, Erkki  
APPLICANT: Pasqualini, Renata  
TITLE OF INVENTION: Molecules That Home to a Selected Organ In Vivo  
NUMBER OF SEQUENCES: 44  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell & Flores LLP  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: United States  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/226,985  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/526,710

FILING DATE: 11-SEP-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/813,273

FILING DATE: 10-MAR-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/862,855

FILING DATE: 23-MAY-1997

ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A.

REGISTRATION NUMBER: 31,815

REFERENCE/DOCKET NUMBER: P-LJ 3423

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 535-9001

TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 7 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-09-226-985-6

Query Match 100.0%; Score 43; DB 3; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2.5e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY 1 CGVRLGC 7

DB 1 CGVRLGC 7

RESULT 4

US-09-227-906-6

Sequence 6, Application US/09227906

Patent No. 6306365

GENERAL INFORMATION:

APPLICANT: Ruostahhti, Erkki

APPLICANT: Pasqualini, Renata

TITLE OF INVENTION: Method of Identifying Molecules That  
TITLE OF INVENTION: Home to a Selected Organ In Vivo

NUMBER OF SEQUENCES: 44

CORRESPONDENCE ADDRESS:

ADDRESS: Campbell & Flores LLP

STREET: 4370 La Jolla Village Drive, Suite 700

CITY: San Diego

STATE: California

COUNTRY: United States

ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/227,906

FILING DATE:

```

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/526,710
FILING DATE: 11-SEP-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/813,273
FILING DATE: 10-MAR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/862,855
FILING DATE: 23-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 3424
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 7 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-227-906-6
Query Match 100.0%; Score 43; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CGVRLGC 7
DB 1 CGVRLGC 7

```

```

RESULT 5
US-07-841-651-4
Sequence 4, Application US/07841651
Patent No. 5410031
GENERAL INFORMATION:
APPLICANT: Rajor, Ana M
APPLICANT: Wright, Ernest M
TITLE OF INVENTION: Cloning and Functional Expression of a
Mammalian Na+/Nucleoside Cotransporter: A Member of
the
TITLE OF INVENTION: SGLT Family
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Sheldon & Mak
STREET: 225 South Lake Avenue, Ninth Floor
CITY: Pasadena
STATE: California
COUNTRY: USA
ZIP: 91101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

```

SOFTWARE: Patentin Release #1.0, Version #1.25  
/  
/ CURRENT APPLICATION DATA:  
/ APPLICATION NUMBER: US/07/841,651  
/ FILING DATE: 19920224  
/ CLASSIFICATION: 435  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: Mandel, Saralynn  
/ REGISTRATION NUMBER: 31,853  
/ REFERENCE/DOCKET NUMBER: 8772  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: (818) 796-4000  
/ TELEFAX: (818) 795-6321  
/ INFORMATION FOR SEQ ID NO: 4:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 662 amino acids  
/ TYPE: AMINO ACID  
/ TOPOLOGY: linear  
/ MOLECULE TYPE: protein  
/ HYPOTHETICAL: NO  
/ ORIGINAL SOURCE:  
/ ORGANISM: Oryctolagus cuniculus  
US-07-841-651-4

Query Match 83.7%; Score 36; DB 1; Length 662;  
Best Local Similarity 71.4%; Pred. No. 77;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CGVRLGC 7  
DB 355 CGTRVGC 361

RESULT 6

US-09-403-089A-8  
/ Sequence 8, Application US/09403089A  
/ Patent No. 6429286  
/ GENERAL INFORMATION:  
/ APPLICANT: SUGIMURA, Kazuhisa  
/ TITLE OF INVENTION: Immunoregulatory Molecules and Process for Preparing  
/ the Same  
/ FILE REFERENCE: 0020-4637P  
/ CURRENT APPLICATION NUMBER: US/09/403,089A  
/ CURRENT FILING DATE: 1999-10-15  
/ PRIOR APPLICATION NUMBER: PCT/JP97/02540  
/ PRIOR FILING DATE: 1997-07-23  
/ PRIOR APPLICATION NUMBER: JP 9/115303  
/ PRIOR FILING DATE: 1997-10-15  
/ NUMBER OF SEQ ID NOS: 8  
/ SOFTWARE: Patentin version 3.0  
/ SEQ ID NO 8  
/ LENGTH: 15  
/ TYPE: PRT  
/ ORGANISM: Artificial  
/ FEATURE:  
/ OTHER INFORMATION: F6 amino acid sequence motif from phage random peptide  
US-09-403-089A-8  
Library

Query Match 79.1%; Score 34; DB 4; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 5;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GVRGCG 7  
 DB 3 GVRGCG 8

RESULT 7

US-09-199-637A-415

Sequence 415, Application US/09199637A  
 Patent No. 6355411

GENERAL INFORMATION:

APPLICANT: Ausubel, Frederick

APPLICANT: Goodman, Howard M.

APPLICANT: Rahme, Laurence G.

APPLICANT: Mahajan-Miklos, Shalina

APPLICANT: Tan, Man-Wah

APPLICANT: Cao, Hui

APPLICANT: Drenkard, Eliana

APPLICANT: Tsongalis, John

TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID

TITLE OF INVENTION: SEQUENCES AND USES THEREOF

FILE REFERENCE: 00786/361002

CURRENT APPLICATION NUMBER: US/09/199,637A

CURRENT FILING DATE: 1998-11-25

PRIOR APPLICATION NUMBER: 60/066,517

PRIOR FILING DATE: 1997-11-25

NUMBER OF SEQ ID NOS: 437

SOFTWARE: FASTSEQ for Windows Version 4.0

SEQ ID NO 415

LENGTH: 219

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-199-637A-415

Query Match 79.1%; Score 34; DB 4; Length 219;  
 Best Local Similarity 85.7%; Pred. No. 61;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVRGCG 7  
 DB 63 GVRGCG 69

RESULT 8

US-09-252-991A-19859

Sequence 19859, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

:/ CURRENT APPLICATION NUMBER: US/09/252,991A  
:/  
:/ CURRENT FILING DATE: 1999-02-18  
:/ PRIOR APPLICATION NUMBER: US 60/074,788  
:/ PRIOR FILING DATE: 1998-02-18  
:/ PRIOR APPLICATION NUMBER: US 60/094,190  
:/ PRIOR FILING DATE: 1998-07-27  
:/ NUMBER OF SEQ ID NOS: 33142  
:/ SEQ ID NO 19859  
:/ LENGTH: 229  
:/ TYPE: PRT  
:/ ORGANISM: pseudomonas aeruginosa  
US-09-252-991A-19859

Query Match 79.1%; Score 34; DB 4; Length 229;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GVRLG 7  
DB 148 GVRLG 153

RESULT 9

US-08-851-843A-178  
Sequence 178, Application US/08851843A  
Patent No. 6093809

:/ GENERAL INFORMATION:

:/ APPLICANT: Cech, Thomas R.

:/ APPLICANT: Lingner, Joachim

:/ APPLICANT: Nakamura, Toru

:/ APPLICANT: Chapman, Karen B.

:/ APPLICANT: Morin, Gregg B.

:/ APPLICANT: Harley, Calvin

:/ APPLICANT: Andrews, William H.

:/ TITLE OF INVENTION: No. 6093809el Telomerase

:/ NUMBER OF SEQUENCES: 225

:/ CORRESPONDENCE ADDRESS:

:/ ADDRESSSEE: Townsend and Townsend and Crew LLP

:/ STREET: Two Embarcadero Center, 8th Floor

:/ CITY: San Francisco

:/ STATE: California

:/ COUNTRY: United States of America

:/ ZIP: 94111

:/ COMPUTER READABLE FORM:

:/ MEDIUM TYPE: Floppy disk

:/ COMPUTER: IBM PC compatible

:/ OPERATING SYSTEM: PC-DOS/MS-DOS

:/ SOFTWARE: Patentin Release #1.0, Version #1.30

:/ CURRENT APPLICATION DATA:

:/ APPLICATION NUMBER: US/08/851,843A

:/ FILING DATE: 06-MAY-1997

:/ CLASSIFICATION:

:/ PRIOR APPLICATION DATA:

:/ APPLICATION NUMBER: US 08/846,017

:/ FILING DATE: 25-APR-1997

:/ CLASSIFICATION:

:/ PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002930US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 178:
SEQUENCE CHARACTERISTICS:
LENGTH: 35 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-851-843A-178

```

```

Query Match 76.7%; Score 33; DB 3; Length 35;
Best Local Similarity 71.4%; Pred. No. 16;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 CGVRLGC 7
DB 2 CGTALGC 8

```

```

RESULT 10
US-08-974-549A-297
Sequence 297, Application US/08974549A
Patent No. 6166178
GENERAL INFORMATION:

```

```

APPLICANT: Cech, Thomas R.
APPLICANT: Lingner, Joachim
APPLICANT: Nakamura, Toru
APPLICANT: Chapman, Karen B.
APPLICANT: Morin, Gregg B.
APPLICANT: Harley, Calvin B.
APPLICANT: Andrews, William H.
TITLE OF INVENTION: Human Telomerase Catalytic Subunit
NUMBER OF SEQUENCES: 727
CORRESPONDENCE ADDRESS:
ADDRESS: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

```



SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/974,549A

FILING DATE: 19-NOV-1997

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/724,643

FILING DATE: 01-OCT-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/844,419

FILING DATE: 18-APR-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/846,017

FILING DATE: 25-APR-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/851,843

FILING DATE: 06-MAY-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/854,050

FILING DATE: 09-MAY-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/911,312

FILING DATE: 14-AUG-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/912,951

FILING DATE: 14-AUG-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/915,503

FILING DATE: 14-AUG-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/US97/17618

FILING DATE: 01-OCT-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/US97/17885

FILING DATE: 01-OCT-1997

ATTORNEY/AGENT INFORMATION:

NAME: Apple, Randolph Ted

REGISTRATION NUMBER: 36,429

REFERENCE/DOCKET NUMBER: 015389-002610US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 297:

SEQUENCE CHARACTERISTICS:

LENGTH: 35 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-974-549A-297

Query Match 76.7%; Score 33; DB 3; Length 35;

Best Local Similarity 71.4%; Pred. No. 16;

Matches 5; Conservative 0; Mismatches 2; Indels 0;

Gaps 0;

QY 1 CGVRLGC 7 ||| ||| |||

RESULT 11  
 US-08-854-050-178  
 Sequence 178, Application US/08854050  
 Patent No. 6261836  
 GENERAL INFORMATION:  
 APPLICANT: Cech, Thomas R.  
 APPLICANT: Lingner, Joachim  
 APPLICANT: Nakamura, Toru  
 APPLICANT: Chapman, Karen B.  
 APPLICANT: Morin, Gregg B.  
 APPLICANT: Harley, Calvin  
 APPLICANT: Andrews, William H.  
 TITLE OF INVENTION: No. 6261836el Telomerase  
 NUMBER OF SEQUENCES: 225  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Townsend and Townsend and Crew LLP  
 STREET: Two Embarcadero Center, 8th Floor  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: United States of America  
 ZIP: 94111  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/854,050  
 FILING DATE: 09-MAY-1997  
 CLASSIFICATION: 536  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/851,843  
 FILING DATE: 06-MAY-1997  
 CLASSIFICATION: 536  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/846,017  
 FILING DATE: 25-APR-1997  
 CLASSIFICATION: 536  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/844,419  
 FILING DATE: 18-APR-1997  
 CLASSIFICATION: 536  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/724,643  
 FILING DATE: 01-OCT-1996  
 CLASSIFICATION: 536  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Apple, Randolph T.  
 REGISTRATION NUMBER: 36,429  
 REFERENCE/DOCKET NUMBER: 015389-002930US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 576-0200  
 TELEFAX: (415) 576-0300  
 INFORMATION FOR SEQ ID NO: 178:

SEQUENCE CHARACTERISTICS:  
LENGTH: 35 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-854-050-178

Query Match 76.7%; Score 33; DB 3; Length 35;  
Best Local Similarity 71.4%; Pred. No. 16;  
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GVRIGC 7  
DB 2 CGTALGC 8

RESULT 12

US-09-430-323-178  
Sequence 178, Application US/09430323  
Patent No. 6309867  
GENERAL INFORMATION:

APPLICANT: Cech, Thomas R.  
Lingner, Joachim  
Nakamura, Toru  
Chapman, Karen B.  
Mortin, Gregg B.  
Harley, Calvin  
Andrews, William H.  
TITLE OF INVENTION: No. 6309867el Telomerase  
NUMBER OF SEQUENCES: 225  
CORRESPONDENCE ADDRESS:  
ADDRESSSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94111

COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/430,323  
FILING DATE: 29-Oct-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/854,050  
FILING DATE: 09-MAY-1997  
APPLICATION NUMBER: US 08/851,843  
FILING DATE: 06-MAY-1997  
APPLICATION NUMBER: US 08/846,017  
FILING DATE: 25-APR-1997  
APPLICATION NUMBER: US 08/844,419  
FILING DATE: 18-APR-1997  
APPLICATION NUMBER: US 08/724,643

FILING DATE: 01-OCT-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Apple, Randolph T.  
 REGISTRATION NUMBER: 36,429  
 REFERENCE/DOCKET NUMBER: 015389-002930US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 576-0200  
 TELEFAX: (415) 576-0300  
 INFORMATION FOR SEQ ID NO: 178:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 35 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: <Unknown>  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 SEQUENCE DESCRIPTION: SEQ ID NO: 178:  
 US-09-430-323-178

Query Match 76.7%; Score 33; DB 4; Length 35;  
 Best Local Similarity 71.4%; Pred. No. 16;  
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGVRLGC 7  
 DB 2 CGTALGC 8

RESULT 13  
 US-09-107-532A-6877  
 : Sequence 6877, Application US/09107532A  
 : Patent No. 6583275  
 : GENERAL INFORMATION:  
 : APPLICANT: Lynn A Doucette-Stamm and David Bush  
 : TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
 : ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND  
 : THERAPEUTICS  
 : NUMBER OF SEQUENCES: 7310  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESS: GENOME THERAPEUTICS CORPORATION  
 : STREET: 100 Beaver Street  
 : CITY: Waltham  
 : STATE: Massachusetts  
 : COUNTRY: USA  
 : ZIP: 02354  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: CD-ROM ISO9660  
 : COMPUTER: PC  
 : OPERATING SYSTEM: <Unknown>  
 : SOFTWARE: ASCII  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/09/107,532A  
 : FILING DATE: 30-Jun-1998  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: 60/085,598  
 : FILING DATE: 14 May 1998  
 : APPLICATION NUMBER: 60/051571  
 : FILING DATE: July 2, 1997

Query Match 74.4%; Score 32; DB 3; Length 212;  
 Best Local Similarity 71.4%; Pred. No. 1.3e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

RESULT 14  
 US-08-861-774E-42  
 : Sequence 42, Application US/08861774E  
 : Patent No. 6297007  
 : GENERAL INFORMATION:  
 : APPLICANT: Waters, Barbara  
 : APPLICANT: Miao, Vivian  
 : APPLICANT: Ho, Yap  
 : APPLICANT: Tong, Seow  
 : TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR  
 : TITLE OF INVENTION: BIOACTIVE MOLECULES  
 : FILE REFERENCE: 9993-006  
 : CURRENT APPLICATION NUMBER: US/08/861,774E  
 : CURRENT FILING DATE: 1997-05-22  
 : NUMBER OF SEQ ID NOS: 94  
 : SOFTWARE: PatentIn Ver. 2.0  
 : SEQ ID NO 42  
 : LENGTH: 212  
 : TYPE: PRT  
 : ORGANISM: Leptogium corniculatum  
 : US-08-861-774E-42

QY 1 CQVRLGC 7  
 DB 108 CQVRLIC 114  
 Query Match 76.7%; Score 33; DB 4; Length 327;  
 Best Local Similarity 85.7%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

ATTORNEY/AGENT INFORMATION:  
 : NAME: Arinello, Pamela Deneke  
 : REGISTRATION NUMBER: 40,489  
 : REFERENCE/DOCKET NUMBER: GTC-012  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: (781)893-5007  
 : TELEFAX: (781)893-8277  
 : INFORMATION FOR SEQ ID NO: 6877:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 327 amino acids  
 : TYPE: amino acid  
 : TOPOLOGY: linear  
 : MOLECULE TYPE: protein  
 : HYPOTHEICAL: YES  
 : ORIGINAL SOURCE:  
 : ORGANISM: Enterococcus faecium  
 : FEATURE:  
 : NAME/KEY: misc\_feature  
 : LOCATION: (B) LOCATION 1..327  
 : SEQUENCE DESCRIPTION: SEQ ID NO: 6877:  
 : US-09-107-532A-6877

```

QY      1  CQVRLGC 7
DB      25  CGVFGC 31

RESULT 15
US-09-198-452A-367
: Sequence 367, Application US/09198452A
: Patent No. 6559294
: GENERAL INFORMATION:
: APPLICANT: Griffiths, R.
: TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides,
fragments
: TITLE OF INVENTION: thereof and uses thereof, in particular for the
diagnosis, prevention
: TITLE OF INVENTION: and treatment of infection
: FILE REFERENCE: 9710-003-999
: CURRENT APPLICATION NUMBER: US/09/198,452A
: CURRENT FILING DATE: 1998-11-24
: NUMBER OF SEQ ID NOS: 6849
: SEQ ID NO 367
: LENGTH: 336
: TYPE: PRT
: ORGANISM: Chlamydia pneumoniae
US-09-198-452A-367

Query Match 74.4%; Score 32; DB 4; Length 336;
Best Local Similarity 57.1%; Pred. No. 2e+02;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1  CQVRLGC 7
DB      30  CGITFGC 36

Search completed: November 13, 2003, 09:54:57
Job time : 8.3125 secs

```

OM protein - protein search, using sw model

Run on: November 13, 2003, 10:27:36 ; Search time 7.875 Seconds  
(without alignments)  
37.610 Million cell updates/sec

Title: US-09-228-866-6

Perfect score: 7

Sequence: 1 CGVRLGC 7

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 118358

Minimum DB seq length: 7  
Maximum DB seq length: 21

Post-processing: Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

1: /cgn2\_6/prodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	100.0	7	7	1	US-08-526-710-6	Sequence 6, Appli
2	100.0	7	7	3	US-08-862-855-6	Sequence 6, Appli
3	100.0	7	7	3	US-09-226-985-6	Sequence 6, Appli
4	100.0	7	7	4	US-09-227-906-6	Sequence 6, Appli
5	85.7	6	15	4	US-09-403-089A-8	Sequence 8, Appli
6	57.1	4	7	3	US-09-139-802-127	Sequence 127, App
7	57.1	4	7	4	US-09-659-786-127	Sequence 127, App
8	57.1	4	7	4	US-08-926-914-127	Sequence 127, App
9	57.1	4	8	2	US-08-318-837-30	Sequence 30, Appli
10	57.1	4	8	3	US-08-444-818-535	Sequence 535, App
11	57.1	4	8	3	US-08-444-818-536	Sequence 536, App

12	4	57.1	8	3	US-08-444-818-537	Sequence 537, App
13	4	57.1	8	3	US-08-444-818-538	Sequence 538, App
14	4	57.1	8	3	US-08-444-818-539	Sequence 539, App
15	4	57.1	9	4	US-09-389-956-92	Sequence 92, App
16	4	57.1	10	4	US-09-311-784A-283	Sequence 283, App
17	4	57.1	14	1	US-07-745-382-4	Sequence 4, Appl
18	4	57.1	14	1	US-07-921-848-4	Sequence 4, Appl
19	4	57.1	14	1	US-08-165-301A-4	Sequence 4, Appl
20	4	57.1	14	3	US-08-810-436-4	Sequence 4, Appl
21	4	57.1	14	5	PCT-US93-06751-80	Sequence 80, Appl
22	4	57.1	14	5	PCT-US94-14179-4	Sequence 4, Appl
23	4	57.1	16	1	US-08-036-555B-121	Sequence 121, App
24	4	57.1	16	1	US-08-469-569-121	Sequence 121, App
25	4	57.1	16	1	US-07-942-245-118	Sequence 118, App
26	4	57.1	16	1	US-08-077-797A-44	Sequence 44, Appl
27	4	57.1	16	1	US-08-249-322A-121	Sequence 121, App
28	4	57.1	16	1	US-08-469-526A-121	Sequence 121, App
29	4	57.1	16	2	US-08-734-591A-121	Sequence 121, App
30	4	57.1	16	2	US-08-469-660-121	Sequence 121, App
31	4	57.1	16	3	US-08-470-335-121	Sequence 121, App
32	4	57.1	16	3	US-08-735-021-121	Sequence 121, App
33	4	57.1	16	3	US-08-734-664A-121	Sequence 121, App
34	4	57.1	16	3	US-08-470-339-121	Sequence 121, App
35	4	57.1	16	4	US-08-467-602-121	Sequence 121, App
36	4	57.1	16	5	PCT-US94-01238-44	Sequence 44, Appl
37	4	57.1	16	5	PCT-US94-05083C-117	Sequence 117, App
38	4	57.1	16	5	PCT-US95-06846A-121	Sequence 121, App
39	4	57.1	18	1	US-08-102-738-1	Sequence 1, Appl
40	4	57.1	20	2	US-07-894-063A-10	Sequence 10, Appl
41	4	57.1	20	2	US-07-894-063A-11	Sequence 11, Appl
42	3	42.9	7	1	US-07-634-641-9	Sequence 9, Appl
43	3	42.9	7	1	US-07-634-641-12	Sequence 12, Appl
44	3	42.9	7	1	US-07-791-213D-34	Sequence 34, Appl
45	3	42.9	7	1	US-08-366-276-6	Sequence 6, Appl

# ALIGNMENTS

RESULT 1  
US-08-526-710-6  
: Sequence 6, Application US/08526710  
: Patent No. 5622699  
: GENERAL INFORMATION:  
: APPLICANT: Ruostiahti, Erkki  
: APPLICANT: Pasqualini, Renata  
: TITLE OF INVENTION: Method of Identifying Molecules That  
: TITLE OF INVENTION: Home to a Selected Organ In Vivo  
: NUMBER OF SEQUENCES: 44  
: CORRESPONDENCE ADDRESS:  
: ADDRESS: Campbell and Flores  
: STREET: 4370 La Jolla Village Drive, Suite 700  
: CITY: San Diego  
: STATE: California  
: COUNTRY: United States  
: ZIP: 92122  
: COMPUTER READABLE FORM:



MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/526,710  
 FILING DATE: 11-SEP-1995  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Campbell, Cathryn A.  
 REGISTRATION NUMBER: 31,815  
 REFERENCE/DOCKET NUMBER: P-LJ 1779  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (619) 535-9001  
 TELEFAX: (619) 535-8949  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 7 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 US-08-526-710-6  
 Query Match 100.0%; Score 7; DB 1; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 2.5e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0;  
 Gaps 0;

QY 1 CGVRLGC 7  
 DB 1 CGVRLGC 7

RESULT 2  
 US-08-862-855-6  
 Sequence 6, Application US/08862855  
 Patent No. 6068829  
 GENERAL INFORMATION:  
 APPLICANT: Rnoshahit, Erkki  
 APPLICANT: Pasqualini, Renata  
 TITLE OF INVENTION: Method of Identifying Molecules That  
 TITLE OF INVENTION: Home to a Selected Organ In Vivo  
 NUMBER OF SEQUENCES: 44  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Campbell & Flores LLP  
 STREET: 4370 La Jolla Village Drive, Suite 700  
 CITY: San Diego  
 STATE: California  
 COUNTRY: United States  
 ZIP: 92122  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/862,855  
 FILING DATE:

```

CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/526,710
FILING DATE: 11-SEP-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/813,273
FILING DATE: 10-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 2621
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 7 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-862-855-6

Query Match 100.0%; Score 7; DB 3; Length 7;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CGVRLGC 7
DB 1 CGVRLGC 7

```

```

RESULT 3
US-09-226-985-6
Sequence 6, Application US/09226985
Patent No. 6296832
GENERAL INFORMATION:
APPLICANT: Rusostahlt, Erkki
APPLICANT: Pasqualini, Renata
TITLE OF INVENTION: Molecules That Home to a Selected Organ In Vivo
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESS: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: United States
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/226,985
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:

```